Stony Brook University

The Impact of DNA Transposable Elements on Allelic Diversity in Maize

A Dissertation Presented

by

Jong-Jin Han

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

December 2012

Stony Brook University

The Graduate School

Jong-Jin Han

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Robert Martienssen – Dissertation Advisor
Professor, Cold Spring Harbor Laboratory

Vitaly Citovsky - Chairperson of Defense
Professor, Stony Brook University

David Jackson
Professor, Cold Spring Harbor Laboratory

Alea Mills
Professor, Cold Spring Harbor Laboratory

Thomas Brutnell
Professor, Donald Danforth Plant Science Center

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

The Impact of DNA Transposable Elements on Allelic Diversity in Maize

by

Jong-Jin Han

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

2012

Transposable elements (TE) are indispensible to understand the evolution of genes and genomes in almost all organisms. DNA transposable elements, class 2 elements, have become a source of material for coordination of eukaryotic gene regulatory systems and for chromosomal reconstruction. In particular, mobility of DNA transposons has been used for insertional mutagenesis to generate new alleles in plants and animals. The well-known *MuDR*/*Mu* system in maize has been adopted for saturation mutagenesis and allowed for the production of myriads of novel mutant alleles in the Maize-Targeted-Mutagenesis (MTM) collection. However, the potential value of this collection has not been properly realized due to the lack of mapped *Mu*-insertion alleles. Here, I show that the position of newly transmitted germinal insertions in the genome can be identified on a large scale via next-generation sequencing technology coupled with a GenomeWalker PCR strategy. I found that more than 100 *Mu* elements per plant transpose mostly into hypomethylated genic regions.

In a parallel study, I show that a 1.8-Mb chromosomal inversion, mediated by a novel *Mu-like* element, is responsible for the classical *Tunicate1* mutation in maize. This is because the inversion causes a fusion between *Zmm19* MADS box gene 5' regulatory region with the 3' end of another gene, GRMZM2G006297. Interestingly, I found a subsequent regional duplication after the chromosomal rearrangement. Sequencing of the duplication break point suggested that CACTA and MITE transposons, which are another classes of DNA transposons in maize, may be involved in this regional duplication, which results in dosage-dependent upregulation of both genes. Taken together, transposon-mediated chromosomal rearrangement and subsequent regional duplication at *Tu1* can influence not only rearrangement of a chromosomal segment but also transcriptional regulatory networks of adjacent genes.

Table of Contents

List of Figures and Tables

List of Abbreviations

A	adenine
A1	anthocyanin1; aleurone pigmentation gene in maize
A2	anthocyanin2; aleurone pigmentation gene in maize
Ac	*activator* locus, DNA transposon
AGL24	*AGAMOUS-LIKE 24*
AP	aspartic proteinase
ATP	ATPase
BM1	*barley MADS1*
bp	base-pair
C	cytosine
CACTA	*Enhancer*/*Suppressor-mutator*-like DNA transposon
Cg1	*Corngrass1*
DDE	arspartate, arspartate, glutamate
DDM1	*Decrease in DNA Methylation1*
diRNA	DSB-induced small RNAs
DIRS	*Dictyostelium* intermediate repeat sequence
DNA	Deoxyribonucleic acid
Ds	*dissociation*
DSN	duplex-specific nuclease
dSpm	defective *Suppressor-mutator* non-autonomous DNA transposon
dsRNA	double-stranded RNA
EN	endonuclease
En	*Enhancer* autonomous DNA transposon

En/Spm	*Enhancer/Suppressor-mutator* DNA transposon
FST	flanking sequence tags
hAT	*hobo/Ac/Tam3*
Hel	helicase
I	*Inhibitor* non-autonomous DNA transposon
I/dSpm	*Inhibitor/defective Suppressor-mutator* DNA transposon
INT	integrase
IS	insertion sequences
Kn1	*knotted1*
LINE	long interspersed nuclear elements
LTR	long terminal repeats
MADS	MCM1/AGAMOUS/DEFICIENS/SRF
MET1	*methyltrasferase1*
MITE	miniature inverted-repeat transposable element
Dnmt1	*DNA* (*cytosine-5*) *methyltransferase1*
Mb	million basepair
mRNA	messenger RNA
MTM	Maize-Targeted-Mutagenesis
Mop1-1	*mediator of paramutation1-1*
Mu	*Mutator*
MuI	*Mu* Inhibitor
MULE	*Mutator-like* transposable element
Non-LTR	non- long terminal repeat

PLE	*Penelope*-like elements
POLB	polymerase B
PRO	protease
PTGS	post-transcriptional gene silencing
qRT-PCR	quantitative reverse-transcription polymerase chain reaction
ra1	*ramosa1*
ra2	*ramosa2*
ra3	*ramosa3*
RC	rolling-circle
RdDM	RNA-dependent DNA methylation
Rdr2	*RNA-dependent RNA polymerase2*
Rep	replication initiator
RH	RNase H
RNA	Ribonucleic acid
RNAi	RNA interference
RPA	replication protein A
RT	reverse transcriptase
SINE	short interspersed nuclear elements
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
Spm	*Suppressor-mutator* autonomous DNA transposon
SVP	*SHORT VEGETATIVE PHASE*
TE	transposable elements
TIR	terminal inverted repeat

TNPA	transposase A
TPase	transposase
Tp1	*Teopod1*
Tp2	*Teopod2*
ts1	*tasselseed1*
ts2	*tasselseed2*
ts4	*tasselseed4*
Ts5	*tasselseed5*
Ts6	*tasselseed6*
TSD	target site duplication
Tu1	*Tunicate1*
Tu1-d	*Tunicate1-d*
Tu1-l	*Tunicate1-l*
Tu1-md	*Tunicate1-md*
Tu1-rec	*Tunicate1 recombinant*
UGI	unique germinal insertion
YR	tyrosine recombinase

Acknowledgments

I would like to thank my Thesis Committee Members: Dr. David Jackson, Dr. Vitaly Citovsky, Dr. Alea Mills, and Dr. Thomas Brutnell for their comments and suggestions on my work. The constructive criticism and insightful comments have been crucial to carry on my research to the right way throughout the years.

I also thank the past and present members of the "team planta" of the Martienssen lab: R. Keith Slotkin, Rebecca Schwab, Milos Tanurdzic, Eyal Gruntman, Patrick Finigan, Michael Regulski, Joseph Calarco, Kate Creasey, Fred Van Ex, Yannick Jacob, Almudena Morales, Evan Ernst, Mark Donoghue, Filipe Borges, Charles Underwood, and Seung Cho Lee as well as members of the "team pombe": Mike Zaratiegui, Danille Irvine, Klavs Hansen, Anna Kloc, An-Yung Chang, Sarahjane Locke, Stephane Castel, Andrea Schorn, Jie Ren, Benjamin Roche, Hyun Soo Kim, and Atsushi Shimada. Together, we have encouraged one another to go through obstacles that we encounter every day. Without them, I can't be who I am. It has been a great privilege to work with them.

I would like to thank my Ph.D. advisor, Dr. Robert Martienssen for providing me with an opportunity to be his student and work on intriguing maize projects. It has been a great pleasure to learn from him. Every time with Rob has become valuable to me.

Finally, I deliver my gratitude to my family who is on my side all the time. Their constant support and encouragement throughout my graduate study have brought me to a state of finishing my graduate course. I do believe that all that I have learned and experienced is a firm foundation that I move forward to develop my career.

Chapter I: Introduction

The discovery and characterization of transposable elements (TE) have led so-called 'selfish junk DNA' to become indispensible elements as to understand genetic and genomic evolution of almost all organisms. Barbara McClintock discovered "jumping genes" or maize DNA transposons in the 1940s, her pioneering work, whose value was highly honored by the Nobel Prize award in 1983, was not at first appreciated by the contemporary scientific community. Her work on unstable alleles in maize revealed the origin of the genomic instability, mediated by the transposition of DNA elements, known as the *Activator* (*Ac*)/*Dissociation* (*Ds*) system (McClintock, 1950). McClintock referred to transposons as "controlling elements' because she recognized that insertion of a transposon at or near the locus of a known gene modifies the action of the gene and even changes phenotypic expression (McClintock, 1956). In the 1970s, it was revealed that insertion sequences (IS) of bacterial DNA sequences are able to transpose (Shapiro and MacHattie, 1979; Shapiro, 1969). Recent advances in genome sequencing have led to a revolution in the understanding of various classes of TEs, which comprise a large fraction of most genomes including fungi, plants, insects, animals, and human (Adams et al., 2000; de Jong et al., 2001; Matsumoto et al., 2005; Szustakowski and Consor, 2001). Their proportions in the genomes are variable depending on the species; for example, about 45% of the human genome, 50-80% of some grass genomes, 3-20% in fungi and 3-45% in metazoans (Bennetzen, 2005; Daboussi and Capy, 2003; Hua-Van et al., 2005; Meyers et al., 2001; Sanmiguel and Bennetzen, 1998; Schnable et al., 2009b). Given the dynamic nature of transposition mechanisms, diversity,

turnover rate etc., these elements have been a critical part in genome evolution (Kidwell and Lisch, 2001; Wessler, 2006)

General characteristics of TEs

Generally, transposons consist of two component systems: 1) an autonomous element encodes proteins necessary for transposition and 2) one or more non-autonomous elements that do not encode transposition-related proteins yet fully rely on the enzymatic activity of transposases for excision, transposition, and insertion. Transposases recognize *cis* elements of the non-autonomous elements in the same cell and direct the mobility of the non-autonomous elements. The site of TE integration is often associate with target site duplications (TSD), which are short direct repeats generated by staggered double strand breaks at the site of integration (Wessler, 2006).

TEs are broadly classified into two groups based on the nature of the transposition intermediate: class 1 elements (RNA-mediated) and class 2 elements (DNA-mediated). Class 1 elements or retrotransposons transpose via an RNA intermediate. The expressed messenger RNA (mRNA) is reverse transcribed by an error-prone reverse transcriptase in the host cell to integrate double-stranded cDNA into the host genome. Autonomous elements encoding reverse transcriptase (RT) and integrase (INT) use their own machinery for reverse transcription and integration processes. Class 2 elements or DNA transposons are transposed by copying their genomic DNA from one chromosomal location to another site in the genome ("cut and paste") without involvement of an RNA intermediate (Jurka et al., 2007; Wessler, 2006). Clearly, TEs using DNA or RNA intermediates are both capable of being amplified and diversified through molecular modifications so that host genome of

organisms has experienced proliferation of TE copy number and diversification of types over evolutionary times.

Class 1 elements

Based on transposition mechanism and structure, Class 1 retroelements can be divided into two groups; LTR (long terminal repeats) retrotransposons and non-LTR retrotransposons (Finnegan, 1989). Autonomous LTR retrotransposons have at least two structural genes; the *gag* gene encoding a capsid-like proten and the *pol* gene encoding a polyprotein that includes aspartic proteinase (AP), reverse transcriptase (RT), RNase H (RH), a DDE (aspartate/aspartate/glutamate) integrase (INT) flanked by LTRs in direct orientation (Wicker et al., 2007). On the other hand, non-autonomous elements lose most of the coding sequence of the autonomous elements so that their internal region is highly variable. The LTR retrotransposons are less abundant in animals, but highly distributed in plants. Grass genomes, in particular, can contain thousands of families predominantly of LTR retrotransposons comprising the majority of their genomes (Haberer et al., 2005; Matsumoto et al., 2005; Messing et al., 2004). In addition to the classical class 1 TEs, two new groups have recently been defined on the basis of integration mechanism; *DIRS*-like elements, which integrate using tyrosine recombinase (YR) instead of integrase (INT), and *Penelope*-like elements (PLEs), which contain genes encoding only reverse transcriptase (RT) and endonuclease (EN) without the *gag* and *pol* genes (Evgen'ev and Arkhipova, 2005; Goodwin and Poulter, 2001, 2004).

Non-LTR retrotransposons are sub-classified into long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Both LINEs and SINEs have

internal RNA polymerase III promoters and terminate by a simple sequence repeat (polyadenylated region) along with TSDs at the insertion sites (Wessler, 2006). Unlike LINEs, SINEs are non-autonomous. Therefore, the autonomous LINEs are necessary for the non-autonomous SINEs to transpose in a *trans*-acting manner (Dewannieux et al., 2003). LINEs vary in prevalence and diversity in eukaryotes; predominant over the LTR retrotransposons in many animals, this class but relatively rare in plants. SINEs are rarely found in plants whereas in humans the *Alu* element at least 500,000 copies in the human genome (Messing et al., 2004; Rowold and Herrera, 2000). However, individual TE activity largely varies from species to species.

Class 2 elements

Eukaryotic DNA transposons are divided into three classes on the basis of transposition mechanism, sequence similarity, and structural relationships; 1) the classic 'cut-and-paste' transposons, 2) *helitrons* using rolling-circle replication, and 3) *Mavericks* using a self-encoded DNA polymerase for replication (Wicker et al., 2007). All cut-and-paste transposons are flanked by 8-13bp TSD, TIRs and encode a single gene for transposase. An enzymatic excision process by transposase is a critical step of the cut-and-paste mechanism. *helitrons* have short conserved terminal motifs and autonomous copies encode a Rep/Helicase, while *Mavericks* have long TIRs and are large transposons that encode multiple proteins including a B-type DNA polymerase. Both *Helitrons* and *Mavericks* move around the genome via a replicative, copy-and-paste process instead of the cut-and-paste mechanism (Kapitonov and Jurka, 2001, 2006, 2007; Pritham et al., 2007).

Subclass 1, which is cut-and paste DNA transposons, has been assigned into super-families (Feschotte and Pritham, 2007; Jurka et al., 2007). 9-13 known superfamilies are classified by the size and sequence of both TSDs and TIRs (Feschotte and Pritham, 2007; Wicker et al., 2007). The 'cut' and 'paste' reactions are catalyzed by a transposase enzyme that recognizes TIRs and introduces DNA nicks at both ends. The number of these TEs can be multiplied by either transposing during chromosome replication from a replicated region to a not-yet replicated position or gap repair mechanism using sister chromatid (Greenblatt and Brink, 1962; Nassif et al., 1994; Wicker et al., 2007). Among these superfamilies, *Mutator* (*MuDR/Mu*), *hAT* (*Ac/Ds*), *CACTA* (*Em/I or Spm/dSpm*) familes have been studied in maize and are known to insert preferentially into genes and low-copy number DNA, which are relatively hypomethylated (Messing et al., 2004).

Mutator

The diverse Mutator superfamily commonly contains long conserved ~220 bp TIRs, which are distinguished from TIRs of other DNA transposons by their long length. However, each class of *Mu* elements has very diverse and unique internal sequences (Bennetzen, 1996). The autonomous element of this family is the self-replicating *MuDR* elements, which regulates mobile activity of non-autonomous *Mu* elements (Chomet et al., 1991; Hershberger et al., 1991). Interestingly, *Mutator-like* transposable elements (MULEs) are found in higher plants such as maize, rice, and *Arabidopsis* and often contain fragments of cellular genes in their internal region surrounded by long TIRs (Jiang et al., 2004; Talbert and Chandler, 1988; Turcotte et al., 2001; Yu et al., 2000). These chimeric elements are called Pack-MULEs that are non-autonomous DNA TEs belonging to *Mutator* superfamily (Jiang et al., 2004). The non-autonomous *Mu* elements are not just deletion derivatives of

MuDR but more likely a fusion element resulting from a duplicated gene fragment flanked by *Mu* TIRs (Lisch, 2002). The insertion generates 9 bp TSDs directly flaking *Mu* TIRs. *Mu* elements have a preference for insertion into or near genes that are unlinked each other over all the maize chromosomes (Lisch et al., 1995; Raizada et al., 2001). *Mu* elements are characterized by high copy numbers, a high rate of germinal transposition (10^{-3} to 10^{-4} per locus), and a low germinal reversion frequency (10^{-5} tagged allele per generation), so that *Mu* insertion mutagenesis has been used as an important tool for a targeted gene-tagging strategy (Alleman and Freeling, 1986; Bennetzen et al., 1993; Candela and Hake, 2008; Chandler and Hardeman, 1992; Lisch, 2002; Lisch et al., 1995; Liu et al., 2009; Rudenko and Walbot, 2001). Taken together, *Mu* elements are ideally suited for genome-wide saturation mutagenesis screens.

To date, four research groups have utilized *Mu* elements for the high-throughput induction of mutant alleles over all ten chromosomes in maize. Firstly, the *RescueMu* transposon collection was generated with a transgenic *Mu1* element that contains a bacterial plasmid for plasmid rescue of flanking sequences. 14,887 non-redundant flanking sequence tags (FSTs) were identified from the *RescueMu* collection (Fernandes et al., 2004). Secondly, the *UniformMu* population was generated by introgressing an active *Mu* into the color-converted W22 inbred to keep mutant alleles in a defined inbred background. Flanking sequences were identified using a modified *Mu* TAIL-PCR and 1,737 non-redundant UniformMu FSTs have been isolated (Settles et al., 2007). Sine then, more FSTs have been characterized and shared with maize community in MaizeGDB.org genome browser. Thirdly, 43,776 individual lines containing stabilized *Mu* insertions from the Maize Targeted Mutagenesis (MTM) population were derived from crosses between active *Mu*

lines and a *Mu-inhibitor* line to transcriptionally suppress *Mu* activity during development of the F1 generation. To isolate FSTs using reverse genetics screens, PCR screens were applied using pooled DNA samples from the F1 generation (May et al., 2003). Fourthly, Pioneer's Trait Utility System for Corn (TUSC) population is composed of about 42,000 F1 plants that are screened by PCR (Bensen et al., 1995; May et al., 2003). Recently, over 40,000 non-redundant *Mu* insertion sites were characterized from Robertson's standard *Mu* stocks obtained by crosses between *Mu* active lines and various inbreds and hybrids (Liu et al., 2009). All the identified FSTs from different *Mu* mutant collection have become an invaluable resource for functional genomics. For the maize community, a collective effort from each group has been to identify new *Mu* alleles and create larger pools of FSTs using a high-throughput deep sequencing technology.

Activator

Activator (*Ac*) and *Dissociation* (*Ds*) were the first transposable elements to be discovered (McClintock, 1950). *Ac* is a member of the *hAT* superfamily whose name was derived from three well-described TE families; *hobo* from *Drosophila*, *Ac/Ds* from maize, and *Tam3* from snapdragons (Calvi et al., 1991). The autonomous 4565 bp *Ac* encodes 807-amino acid transposase enzyme that recognizes TIRs and TPase binding sites at 5' and 3' subterminal region of *Ac* itself and the non-autonomous *Ds* elements for *cis*- and t*rans*-activation (Becker and Kunze, 1997; Coupland et al., 1988; Houbaherin et al., 1990). Although a consensus AAACGG binding site in the 5' and 3' 200 bp subterminal repeats was strongly suggested *in vitro*, strong target site consensus sequences for *Ac* and *Ds* have not been obvious identified in maize (Grotewold et al., 1991; Kunze and Starlinger, 1989;

Vollbrecht et al., 2010). Unlike *Mu* elements, most *Ds* elements are derived from *Ac* by internal deletions between short direct duplications (Kunze and Weil, 2002). *Ac* transposase recognizes TIRs of non-autonomous *Ds* elements and/or itself by landing on its binding sites, and tend to mobilize them into other chromosomal locations locally (Cowperthwaite et al., 2002; Dooner and Belachew, 1989; Vollbrecht et al., 2010). This feature is effectively utilized both for regional mutagenesis that saturates nearby genes with insertions from a linked *Ac* and for reconstitutional mutagenesis providing new alleles by reinserting *Ac*/*Ds* at the different site in the same gene (Bai et al., 2007; Moreno et al., 1992). Although relatively low copy number and a biased distribution of insertion sites limit genome-wide saturation mutagenesis, the enrichment of both transposed-*Ac* elements (*tr-Acs*) and transposed-*Ds* elements (*tr-Dss*) have made available the widespread distribution of *Ac* and *Ds* "launchpad" insertions throughout the genome (Kolkman et al., 2005; Vollbrecht et al., 2010). This system has been useful in regional mutagenesis targeting both linked and unlinked positions (Vollbrecht et al., 2010).

Enhancer/Suppressor-Mutator

Enhancer/Suppressor-mutator (*En/Spm*) is a member of the CACTA superfamily. TEs of the CACTA superfamily contain both a transposase and another ORF of unknown function (Wicker et al., 2003). Elements from the CACTA superfamily terminate in the well-conserved 5'-CACTA…TAGTG-3' motif before a 3 bp TSD in plants and a 2 bp TSD in animals (Wicker et al., 2007). The *En/Spm* family has TIRs, subterminal regions containing the *cis* sequences necessary for transposition, and the internal region containing genes encoding *trans*-acting proteins like *Ac/Ds system.* The autonomous *Enhancer* (*En*)

and the non-autonomous element *Inhibitor* (*I*) were named by Peter Peterson and the same elements were named by McClintock as *Suppressor-mutator* (*Spm*) and defective *Spm* (*dSpm*), respectively (McClintock, 1954; Peterson, 1953). The insertion of defective *Spm* (*dSpm*) in mutable alleles of the anthocyanin genes *A1* and *A2* and *En/Spm* insertions *al* locus were also cloned (Masson et al., 1987; McClintock, 1965; O'Reilly et al., 1985). *En/Spm* has been reported with two molecular functions. In the absence of an autonomous *En/Spm*, the non-autonomous *I/dSpm*-inserted mutant alleles decrease but not eliminate their gene expression by splicing the *I/dSpm* element out of the progenitor RNA (McClintock, 1956; Menssen et al., 1990; Raboy et al., 1989). *En/Spm* contains splicing signals that make itself available to be spliced out of the transcripts and to function as a mobile intron (Purugganan and Wessler, 1992). When *En/Spm* is present in the genome, gene activity from the alleles where *En/Spm* is located is completely suppressed, but the non-autonomous *I/dSpm* excises and reinserts elsewhere like the *Mu* activity (McClintock, 1954, 1961). This suppressor function results from TNPA protein encoded by the *En/Spm* element (Grant et al., 1990). In terms of transposon tagging, *En/Spm* tends to transpose more widely than *Ac/Ds,* which transposes relatively short distances along the chromosome.

***Helitrons and Mavericks* (*Polintons*)**

Like other transposons, *Helitron* transposons have been identified in all eukaryotic organisms but their copy numbers are highly variable (Kapitonov and Jurka, 2007). The structural features of *Helitrons* are the rolling-circle (RC) replication initiator (Rep) and DNA helicase (Hel) domains, which are encoded by all autonomous *Helitron* elements. The Rep domain contains three motifs that are involved in endonucleolytic cleavage, DNA

transfer, and ligation activities. The Hel domain has eight motifs that are used for helicase activity (Kapitonov and Jurka, 2001). Interestingly, *Helitrons* insert into AT target sites of the host with 5'-TC and CTRR-3' termini (R stands for A or G) and contain about 20 bp hairpins derived from 10-12 bp of the 3' end instead of TSDs during transposition due to utilization of RC mechanism (Kapitonov and Jurka, 2001). The hairpin palindromic sequences are thought to function as the terminator of rolling-circle replications. Thus, the heterogeneous *helitrons* are different from the classical cut-and-paste DNA transposons in their structure and transposition mechanism so that the short termini and possibly the hairpin are the only structural features that can be applied in identification at present.

Mavericks, also known as *Polintons*, is the third class of DNA transposons characterized by *in silico* computational studies. *Mavericks/Polintons* are the most complex eukaryotic transposons encoding a protein-primed DNA polymerase B (POLB), a retroviral-like integrase (INT), an putative ATPase (ATP), and an adenoviral-like cysteine protease (PRO) with 6 bp TSDs, 0.1-1 kbp TIRs, and 5'-AG and TC-3'termini (Kapitonov and Jurka, 2006). *Mavericks/Polintons* are thought to propagate through protein-primed self-synthesis mechanism by their POLB. When host genome is replicated, the INT excises *Mavericks/Polintons* from the host DNA as extrachromosomal single-stranded DNA. Double-stranded extrachromosomal elements replicated by POLB are integrated back into the host genome by INT (Kapitonov and Jurka, 2006). The high complexity of *Mavericks/Polintons* may limit their expansion in the host genome, which explains the low copy numbers of this family in most genomes.

Chromosomal rearrangement by DNA transposons

Double-strand breaks and repair events induced by DNA transposons actively facilitate large-scale chromosomal rearrangements and change the genome structure in almost all organisms including plants and animals (Bennetzen, 2005; Bourque, 2009; Parisod et al., 2010). McClintock described the chromosome-breaking *Ds* element that has a complex structure in maize. For example, one 2.2 kb element inserted inside and the other placed in inverse orientation, resulting in chromosomal breaks due to two pairs of 5' and 3' ends facing each other over a short distance (Kunze and Weil, 2002). Also, *Ac* and *Ds*, which were not in close proximity but more than 100 kb apart, led to chromosome breaks (Weil and Wessler, 1993; Zhang and Peterson, 2004). Transposase-induced rearrangements can mediate the restructuring of the chromosome between DNA transposable elements. Alternative transposition can take place between two individual transposon copies whose termini are contacted in a partial cut-and-paste activity, resulting in chromosomal deletions, inversions, duplications, and translocations (Gray, 2000; Lim and Simmons, 1994; Zhang and Peterson, 2004). In chapter 4, other evidence of chromosomal rearrangements caused by *Mutator-like* elements is presented. Chromosomal inversion and regional duplication differently manipulate neighboring gene expression that results in a drastic phenotypic alteration.

DNA transposons in gene duplication and exon shuffling

DNA transposons contribute to the evolution of new genes by capturing a complete or incomplete host gene, such as *Mutator*-like elements (MULEs) from maize, rice, Arabidopsis and *Helitron* elements from maize, and CACTA elements from soybean and Antirrhinum (Hoen et al., 2006; Jiang et al., 2004; Morgante et al., 2005; Roccaro et al.,

2007; Talbert and Chandler, 1988; Zabala and Vodkin, 2005). MULEs have been shown to carry fragments of cellular genes and these chimeric elements have been called Pack-MULEs (Jiang et al., 2004). These elements were first isolated in maize and Pack-MULEs are capable of amplifying genes or gene fragments massively. (Lisch, 2005; Talbert and Chandler, 1988). In rice, nearly 3000 Pack-MULEs could have fragments from more than 1000 cellular genes. 22% of the assembled chimeric genes are even transcribed into novel transcripts. Also, small RNA are associated with Pack-MULEs to suppress the expression of their own transcripts and the parental genes which are the source for internal region of MULEs (Hanada et al., 2009). As Pack-MULEs are able to duplicate the parental genes or gene fragments and express them, they may be involved in changing the regulatory network of duplicated genes and be a resource for coding sequences in host genomes. For example, in *Arabidopsis*, MULEs have incorporated host sequences for the protease domain of ubiquitin like protein-specific proteases (ULP) (Hoen et al., 2006).

Ancient *Helitron* transposons evolved by capturing replication protein A (RPA) and DNA helicase (Hel) genes from the host (Kapitonov and Jurka, 2001). Similarly, partial or intact copies of host genes have been found in non-autonomous *Helitrons* in the maize genome (Gupta et al., 2005; Lai et al., 2005; Lal et al., 2003). Highly diverse gene content among the "McC", B73, and Mo17 maize inbred lines suggests the lack of genic colinearity, which can be partially explained by the presence/absence of non-autonomous *Helitrons* containing host genes in these inbred lines (Fu et al., 2001; Lai et al., 2005; Song and Messing, 2003). In two maize inbred lines, 10,000 DNA insertions were identified by *helitron* termini with multiple gene-derived fragments internally and had a structure typical of non-autonomous *helitron*-like transposons. Not only MULEs but also *helitrons* using

distinct rolling-circle replication mechanism have a huge potential to function as exon shuffler (Feschotte and Wessler, 2001; Morgante et al., 2005). In one example, a *helitron* carrying an cytidine deaminase gene has been characterized in maize (Xu and Messing, 2006). But the details of the mechanisms of gene capture by *helitrons* are still unclear: DNA replication errors may allow *helitrons* to capture host gene sequences during repair of double strand breaks caused by DNA transposon excision (Dooner and Weill, 2007). Clearly, Pack-MULEs and *helitrons*, as well as other DNA transposons, have contributed to genome plasticity by generating a constant change of genic and non-genic regions by gene duplication and gene shuffling to some extent.

Alteration of gene expression by transposition of DNA transposons

DNA transposons and retrotransposons can directly influence the regulation of gene function by restructuring sequences in genic regions. Gene expression at the transcriptional level depends on where transposons are inserted. The disruption of coding sequences causes a complete loss of gene function, while perturbation of regulatory elements such as enhancer, silencer, promoter, and untranslated region leads to subtle mis-expression genetically (Kapitonov and Jurka, 2006; Kidwell and Lisch, 1997). However, the tendency of DNA transposons to insert in genic regions can be a driving force to generate allelic diversity in the host genomes. Unlike retrotransposons, DNA transposons can make unstable mutations with reversible phenotypes (by spontaneous excision), or more permanent mutations with footprints caused by imperfect excision (Kidwell and Lisch, 1997; Plasterk, 1991). Thus, DNA transposon excision and insertion can allow for the creation of new alleles and novel modification of regulatory units, giving rise to the increase of allelic diversity.

Epigenetic regulation of DNA transposons

McClintock described some of the first examples of epigenetic silencing as heritable and reversible changes in "phase" of DNA transposons *Spm* and *Ac* (McClintock, 1984). Characterization of the active and inactive phases of *Mutator* supported the idea that activity of *Mu* elements is correlated with methylation of cytosine residues (Chandler and Walbot, 1986). Molecular analyses of *Spm*, *Ac/Ds* and Robertson's *Mutator* has shown *Mu* activity is affected by changes in DNA methylation, especially in the TIRs (Banks et al., 1988; Bennetzen, 1996; Chomet et al., 1987; Martienssen et al., 1990; Martienssen and Baron, 1994). In these cases, hypermethylated TIRs suppress activity of the autonomous element, but active transcription is recovered once methylation is lost. Lippman et al. (2004) revealed that DNA transposons play a major role in formation and maintenance of heterochromatin in which repetitive sequences including CACTA and MULE DNA transposons are enriched in CG and H3K9 methylation.

CG methylation in *Arabidopsis* is mediated by DNA *Methyltrasferase1* (*MET1*), which is a homologue of *DNA* (*cytosine-5*) *methyltransferase1* (*Dnmt1*) in mammals (Aufsatz et al., 2004). Also, the chromatin remodeling SWI2/SNF2 ATPase, *Decrease in DNA Methylation1* (*DDM1*) is required for CG methylation maintenance and a suppressor of transgene silencing (Jeddeloh et al., 1998; Vongs et al., 1993). In *ddm1* mutants, repetitive sequences are hypomethylated and MULEs and *Spm*-like CACTA transposons are reactivated and mobilized into other locations at high frequencies, suggesting that DDM1 epigenetically regulates gene expression and transposon activity (Miura et al., 2001; Singer et al., 2001). Moreover, *modifier of paramutation1* (*mop1*) gene, which is a homolog of *Arabidopsis RNA-dependent RNA polymerase2* (*RDR2*), has been shown to be

epigenetically involved in silencing/reactivation of *Mu* elements in maize as well as paramutation (Lisch et al., 2002; Woodhouse et al., 2006).

When DNA transposons are epigenetically silenced the insertion of the silenced transposon in regulatory regions can bring the adjacent gene under transposon control in both in *Arabidopsis* and in maize (Barkan and Martienssen, 1991; Lippman et al., 2004). These data support the idea that epigenetic modification, especially DNA methylation, of regulatory regions can alter activity of DNA transposons and influence patterns of neighboring gene expression. However, it is still little known what drives epigenetic modifications in DNA transposons and how alterations of methylation level are controlled.

Co-option of DNA transposon sequence

Just as transposons can acquire host genome sequences, the host can acquire transposon sequences as a source of raw material that can be used for the assembly of new genes and functions (Volff, 2006). DNA transposons have useful properties for molecular domestication or exaptation by the host genome (Feschotte and Pritham, 2007). The *Arabidopsis* transcription factors FHY3 and FAR1 which regulate genes in the phytochrome A signaling pathway are structurally similar to the JITA transposase of the maize element *Jittery*, which is a member of *Mutator* transposon superfamily (Hudson et al., 2003; Xu et al., 2004). Although FHY3 and FAR1 have not retained TIRs at their termini, nor transpositional function due to sequence modification over time, they have adopted the function of *Mutator* transposase that regulated transcription of its own gene and other MULEs. Similarly, *MUSTANG* has been also derived from *Mutator* transposase and is not flanked by TIRs in *Arabidopsis* and rice (Cowan et al., 2005). Also, the transposase gene of

hAT elements has been domesticated into the *Gary* gene in grasses, and into the *DAYSLEEPER* gene in *Arabidopsis,* as a DNA-binding factor targeting the upstream of the DNA repair gene *Ku70.* These genes have the DNA-binding domain but have lost sequences for transposition such as the TIRs (Bundock and Hooykaas, 2005).

Summary

For more than half a century, diverse organisms have been studied to understand transposition mechanisms and the impact of the mobile 'controlling gene' on genetic and genomic evolution. DNA transposons have a preference to transpose into genic regions and to control host gene regulation depending on insertional position and epigenetic status. Such unique features of class 2 elements are vitally important to advance the usefulness of DNA transposons for insertional mutagenesis as a critical resource to obtain new alleles. In Chapter 2, I describe how the Maize-targeted-Mutagenesis (MTM) population has great potential as a resource for new alleles, but it has not been efficiently utilized due to technical limitations of reverse genetics. A high-throughput next generation sequencing technology has allowed us to create a sequence-indexed FST database by identifying *Mu* transposon insertion sites and to investigate *Mu* activity on a large scale. Among the many of McClintock's discoveries on *Ac/Ds* transposons, the negative dosage effect between autonomous *Ac* elements has remained mysterious. In Chapter 3, although posttranslational regulation by *Ac* transposase is a controversial hypothesis, we have approached the question with the notion of *Ac* autoregulation via transposon-derived small RNAs. Finally, *Tunicate1* is a classical mutant of maize first described over 200 years ago. Despite its striking phenotype and plausible relatedness to maize domestication, the molecular rearrangement

underlying the dominant *Tu1* gene has remained mysterious. In Chapter 4, we find that a MULE transposon-mediated inversion/duplication accounts for the ectopic expression of this important gene and discuss the implications.

Chapter II: Identification of Sequence-indexed *Mutator* (*Mu*) insertion sites and *Mu* transposition behavior in Maize-Targeted-Mutagenesis (MTM) population

Introduction

Although homologous recombination and RNA interference (RNAi) techniques have been applied to knock out and/or knock down genes in other organisms, these strategies are impractical in maize due to the relatively low rate of homologous recombination and difficulty of transformation (Carpenter and Sabatini, 2004; May et al., 2003). Comprehensive insertional mutagenesis allows for genome-wide functional screens to investigate the association of genes to functions in plants (McCarty et al., 2005). In *Arabidopsis,* sequence-indexed T-DNA and maize transposon-induced insertion lines have been developed to knock out almost every single gene in the genome (Alonso, 2003; Parinov and Sundaresan, 2000). Maize transposons have provided one of the most important tools for gene identification and characterization in plants. Using characteristics of each DNA transposon, low-copy *Ac/Ds* and high-copy *MuDR/Mu* strategies, both of which target genic regions, have been adopted to generate multiple insertion-tagged maize populations (Kolkman et al., 2005; May et al., 2003; Settles, 2005). Whereas *Ac/Ds* elements are inefficient global but efficient regional mutagens because of low copy numbers, low forward mutation rate, and propensity for transposition to the site closely linked to the original copy, *MuDR/Mu* transposons fit in the goal of globally saturation mutagenesis due

to high copy number in an active line and a high forward mutation versus a low germinal revertant frequency (Lisch et al., 1995; Walbot, 2000).

One requirement for *Mu* activity is the presence of abundant *MuDR* transcripts (Chomet et al., 1991). Most new mutations in *Mutator* lines are mediated by transposition of nonautonomous elements whose number are 10 times more abundant than the *MuDR* elements (~100 versus ~10) (Bennetzen, 1996). However, the known estimate of *Mu* copy number is not reliable because *Mu* copy number is variable according to inbred backgrounds where *Mu* active lines are outcrossed and also newly-isolated *Mu* elements that are not cross-hybridized with *Mu1* are not included (Alleman and Freeling, 1986; Bennetzen, 1984). New mutations in *Mutator* are germinally transmitted and associated with a replicative mechanism that doubles *Mutator* copy number of the self-pollinated progenies, compared to their parents (Alleman and Freeling, 1986; Lisch et al., 1995). Also, duplications were detecteced in minimal lines carrying a single *MuDR* element (Lisch et al., 1995). *Mu* elements can maintain their copy number following multiple outcrosses to non-*Mutator* lines through this replicative transposition process (Alleman and Freeling, 1986; Walbot and Warren, 1988). Whereas *Mu* elements excise and insert apparently using a cut-and-paste-mechanism late in somatic development, replicative transposition that increase in *Mu* element copy number occurs in the pre-meiotic cells, meiotic cells, and haploid gametophytic cells (Walbot and Rudenko, 2002). Two models have been proposed to explain this phenomenon. One model is the gap-repair model. After the *Mu* element has transposed, the gap at the excision site needs to be repaired. In germinally transmitted cell lineages, the sister chromatid is used for gap repair, resulting in the *Mu* duplication instead of reversion but not in somatic cells (Lisch, 2002). Another model is the

replicative/duplicative model proposing that germinally transmitted lineages take a semi-conservative duplicative transposition mechanism like bacterial transposon *Tn7* (May and Craig, 1996). Since *Mu* insertions are independent in individual progeny, germinal insertions accumulate with preservation of existing *Mu* locations by a replicative transposition.

To reduce the frequency of somatic transpositions, the Maize-Targeted-Mutagenesis (MTM) population was generated by a cross between a *Mu*-active lines and a line carrying a *Mu* inhibitor (*MuI*) line as a dominant repressor of *Mu* (Martienssen and Baron, 1994; May et al., 2003). DNA from 43,776 F1 plant leaves that are *Mu*-inactive were collected into a 2-dimensional grid after harvest and pooled into rows and columns. The F1 progenies were self-fertilized and assigned barcode numbers for the scientific community pursuing a reverse genetics in maize (May and Martienssen, 2003).

To date, it remains unclear how many *Mu* elements are present in *Mu* active lines and how many new insertions are germinally generated every single generation despite of estimate numbers of *Mu* copy number (50 – 100 copies or more) (Bennetzen, 1996). Here, we present the utility of MTM flanking sequence tags (FSTs) for reverse genetics by validating FSTs obtained via a high-throughput Illumina sequencing. Also, molecular behaviors of *Mu* elements in MTM population will be described.

Results

Crosses to generate MTM lines led to diminish somatic insertions that cause false positives in most F1 progenies and stabilize new germinal insertions that occurred during pre-meiotic or meiotic gametophyte development (May et al., 2003). For the crosses, two types of *Mu*-active parents were used: 1) carrying *bz1-Mum9* allele in which *Mu* element is inserted in 2nd exon of *bz1* reporter gene expressed in the aleurone layer of the endosperm and 2) carrying *a1-Mum2* in which *Mu* element is inserted in the *a1* aleurone color gene (Brown and Sundaresan, 1992; Chomet et al., 1991). To reduce somatic insertions and stabilize germinal insertions, these active *Mu* lines were crossed with a line that carries both *Les28*, a dominant lesion mimic mutation, and unknown dominant sectoring factors that suppress *Mu* activity (Martienssen and Baron, 1994). As *Mu* activity was epigenetically repressed by the *Mu* inhibitor in the upper leaves during development of the F1 generation, the *Les28* phenotype depending on *Mu* behavior disappeared and the *Mu*-active spotted seeds also were rarely observed. The dried tissue samples (two upper leaves) were arranged in 48-by-48 plant grids and DNA was extracted from the pooled tissues constituting each column and row. As a result, 96 DNA samples were prepared from each grid of 2,304 (48 X 48) plants. Because each row and column were sampled from separate leaves, a double-sampling strategy allows for separation of germinal insertions in a whole plant from somatic insertions in leaf sector by intersecting its unique column and row address (Das and Martienssen, 1995; May et al., 2003). We used one grid M04 (96 pooled DNA from 48 columns and 48 rows) that were the progeny of *Mu*-active parents carrying *bz1-Mum9* and the *Mu*-inhibitor line for library generation.

Construction of a 4nt-barcoded MTM libraries

A single pooled DNA from column and row is the mixture of 48 individual plants. To maximize representation of unknown flanking DNA sequences of MTM *Mu* insertion sites, we have modified the ligation-mediated GenomeWalker protocol (Clontech laboratories, Inc.) outlined in Figure 2.1A (McCarty et al., 2005; Siebert et al., 1995). The first step of generating GenomeWalker-coupled MTM libraries is to construct pools of adaptor-ligated genomic DNA fragments. For DNA fragmentation, five blunt end restriction enzymes were selected; three methylation-sensitive restriction enzymes (FspI, NruI, SmaI) to enrich *Mu* flanking regions that are mostly hypomethylated and two methylation-insensitive restriction enzymes (DraI, StuI) not to exclude *Mu* insertion sites other than genic regions. For better ligation efficiency, we added 'A' base to the 5' end of the DNA fragments and ligated them with the modified GenomeWalker adaptors with 3' T overhang. In addition, the adaptors was further modified by incorporating six series of 4 nt-barcodes into the 3' end of the long-arm strand and just before 3'T modification (Figure 2.1B). In particular, the adaptors are designed to prevent adaptor-adaptor amplification by blocking extension of the 3' end of the short-arm strand with the amine group. The second step is to amplify only *Mu* flanking sequences selectively using nested adaptor-specific primers and nested universal Mu TIR degenerate primers. For the high-throughput large-scale Illumina sequencing, the nested second PCR primers possess long oligos that include both Iluumina SBS sequencing primer sequence and the sequence complimentary to the immobilized flow cell primer. As a consequence, the purified final MTM libraries are suitable for 76 bp paired-end (PE) read sequencing. Instead of SBS3 Illumina sequencing primer, our custom sequencing primer, same as the nested second adaptor-specific primer without long oligos,

successfully worked for the first run, resulting in recovery of extra 19 nt *Mu* flanking sequences. For the second run, Illumina SBS8 primer was applied. A plot of cluster intensity versus cycle number is shown in Figure 2.2A. The presence of a series of 4 nt barcodes allowed multiplexing to reduce costs. In this study, 6 libraries were able to be sequenced in one lane of a flow cell and 48 libraries in one whole flow cell in Illumina Genome Analyzer II (GAIIx). Two full flow cells were capable of 76 nt PE sequencing of 96 libraries, which are 48 from columns and 48 from rows in one grid.

As a pilot data analysis, we first mapped 12 libraries (C01 to C06 and R01 to R06) to B73 maize reference genome (Maize B73 RefGen_V2). Among 99 and 262 million reads passing filter from one lane (C01-06) and other lane (R01-06), respectively, 99.1% of them contained 4 nt barcodes and almost evenly splitted into 6 libraries. After trimming 5 nt (4nt barcode plus universal 1nt T base) from the 1st end, 71nt maize flanking sequence reads were recovered. On average, 94% and 97% of the barcoded reads had *Mu* TIRs on the other end of the read from libraries C01-06 and R01-06, respectively. This suggests that 93% of 99 million reads from C01-06 have barcode and *Mu* TIR sequences at each end, respectively, and 96% of 262 million reads from R01-06 do two end sequences. From the 2nd ends, almost universal 26 nt *Mu*TIRs with a few single nucleotide polymorphisms (SNPs) among *Mu* family members were removed to collect 50 nt flanking sequence tags (FSTs) containing the exact *Mu* insert position right next to the end of *Mu* TIR. The maize genomic sequence from both ends insures sufficiently unique mapping to the maize genome sequence. The authentic flanking sequence information from two ends was used as a query for mapping against B73 maize genome. As MTM plants are in an unknown genetic background, about 39% of C01-

06 and 61% of R01-06 flanking sequence reads were eventually mapped to a unique location in the maize genome.

Overall distribution of MTM *Mu* target sites

Using the 12 data sets generated from the MTM libraries (from 6 column pools and 6 row pools) analyzed, *Mu* target sites were distributed uniformly across chromosomes, whereas intra-chromosomal distribution is non-uniform, which will be presented in later section. As shown in Figure 2.3, the proportion of *Mu* target sites in all 10 chromosomes is relatively similar among 12 data sets. Number of *Mu* insertion sites in each chromosome is relatively proportional to the size of each chromosome (data not shown). This suggests that insertional preference of *Mu* elements is not regional but global and supports the notion that *MuDR/Mu* system is suitable for saturation mutagenesis in maize (Walbot, 2000). Unexpectedly, we could not detect any *Mu* insertion sites on chromosome 7 to 9 from Columns 2 and 3 data sets as well as chromosome 5 to 9 in Column 4. It is very unlikely that library generation and illumina sequencing procedures selectively exclude these chromosomes in three specific pools and/or *Mu* elements in these pools are specifically hampered to jump into them. It might happen that a humanly mistake caused the exclusion of the corresponding chromosome information during the processing of sequencing data.

Identification of parental insertions (PIs)

Parental insertions are shared in most of progenies so that they are theoretically 48 times more abundant than germinal insertions that are unique to only one plant (Figure 2.2B). To identify the number of insertions present in the parents of the MTM population,

we pulled out the insertion coordinates that were shared between all 12 libraries, which are 6 non-overlapping column libraries and 6 non-overlapping row ones. Surprisingly, 572 parental insertion candidates were identified from this analysis. Based on the previous studies stated in introduction sections, this number is much higher than what it is expected. Furthermore, MTM *Mu* elements most preferentially target 5' end region in which *cis*-regulatory elements are positioned in Figure 2.4. Almost half of *Mu* target sites are 5' end region that is within 1 kb upstream of start codon, consistent with previous studies (May et al., 2003; Settles et al., 2007). However, a large portion of their distribution data were unknown sequences, due to incomplete B73 genome sequence and lack of annotated gene transcript at the given time. As we will describe in discussion section, it is possible that majority of parental insertions would be derived from immovable Pack-MULEs that are preferentially located with 500 bp upstream from 5' termini of genes (Jiang et al., 2011). If flanking regions of PACK-MULEs occupy most of parental insertions, their 5' end preferential insertions would alternatively explain a high ratio of 5' end insertion from the parental insertion candidates. On the other hand, intergenic region occupies less than 20%, which is reasonable, yet it is possible that our strategy to enrich *Mu* flanking region using more methylation-sensitive enzymes may exclude *Mu*-inserted intergenic region to some extent.

To validate parental insertions the PCR validation assays were applied with the locus-specific primer and a MuTIR degenerate primer for four genes, GRMZM2G306079 and GRM2G306079, GRMZM2G438007, and GRM2G110582 (Figure 2.5 and Figure 2.6). We detected that *Mu* parental insertions were transmitted to the F3 progenies of randomly-selected four MTM lines. The PCR assays revealed that sib-crosses of MTM F2 progeny

that carried homozygous *Mu*-inserted parental alleles allowed for inheritance of the allele in all F3 progeny and those of heterozygous *Mu* alleles for being segregated in next generation.

Identification of unique germinal insertions (UGIs) in 36 individual MTM plants

To identify number of UGIs, parental insertions were first subtracted and new germinal insertions were extracted by intersecting coordinates of *Mu* insertion sites from both column and row. Unique insertions can be only shared by cross-checking one column and one row because we used the double-sampling strategy for collecting two upper leaves assigned to column and row in a single MTM plant. If unique insertions are found only either in row or column, they are considered as non-heritable somatic insertions that are false positive. Figure 2.7 shows that about 100 up to the most 280 unique germinal insertion candidates were identified in 36 MTM plants. On average, 162 new germinal insertions per MTM plant were identified so that we identified about 5800 unique germinal insertion candidates from 36 MTM lines. The actual number of new germinal insertions can be fluctuated if further bioinformatic analyses are performed with 96 full sets of sequencing data as well as do additional PCR validation experiments for every singe insertion (which may be impossible due to costs), yet the variation would be expected not to be drastic. As every single insertion is unique, the same insertion site is not present in other 35 lines. In case two individual *Mu* elements are found in the same locus, they are recognized as different alleles because the distance between the insertion sites are at least 20 bp apart.

Biased distribution of unique germinal insertions (UGIs) within chromosome

The coordinates of all UGIs collected from 36 plants were used for plotting numbers of MTM *Mu* insertions per 10 Mb (Million Basepair) in maize chromosome 1 through 4. UGIs are non-uniformly distributed on each of the chromosomes, forming the 'U-shaped' pattern in Figure 2.8A to 2.8D. The frequency of *Mu* insertions exponentially increases as the coordinate of each chromosome move away from centromere(s), indicating that *Mu* elements exhibit a transposition tendency to dip toward centromere(s) but peak toward the arms of chromosomes. This U-shaped pattern resembles a pronounced 'bowl-like' trend observed in another Robertson's *Mu* line, which is associated with meiotic recombination events (Liu et al., 2009). Robertson's *Mu* elements in MTM lines confirm biased distribution of *Mu* insertions that are associated with relative distance from centromeres, gradually hiking toward the ends of each chromosome.

Demethylation of *Mu*-inserted flanking sequence

To investigate the relatedness of DNA methylation level with *Mu*-inserted flanking regions, maize methylation data were examined (Regulski et al., submitted). Interestingly, methylation level of flanking regions where *Mu* targeted is overall lower than hypomethylated genes and further hypomethylated exons in symmetric CG and CHG contexts (Table 2.1). However, there is no significant change in asymmetric CHH context. Recently, (Gent et al., 2012) showed that chromosome-level patterns of CG and CHG methylation increase toward centromeres but decrease toward arms in maize. This suggests that demethylated *Mu* flanking regions may be more frequently found toward the ends of

each chromosome into which *Mu* elements are intensively transposed, indicating that DNA methylation is negatively correlated with frequencies of *Mu* insertions.

Validation of UGIs

UGI candidates were tested in the same way as did PI PCR validation assays. For GRMZM2G88689 and GRM2G106179 genes, F3 progenies were successfully genotyped using a locus-specific forward and reverse primer pair along with MuTIR4 primer in Figure 2.9. *Mu* insertion in 5' UTR of GRMZM2G88689 shows that both *Mu* TIR ends were validated with a FST locus-specific primer pair. Even *Mu*-spanning PCR product, which refers to either heterozygous or homozygous allele, was successfully amplified in progeny number 3, 6, and 7. Further, Figure 2.10 presents four UGI examples whose function is known. Despite some non-specific PCR background, the expected size of PCR amplicons was detected from all the cases that we tested except one. At this point, the failure of one case would be a matter of primer design. If more investigation for UGI validation is carried out by PCR validation assays, the authenticity of *Mu* insertions will be verified more robustly.

Discussion

In this study, we improved a protocol of constructing MTM DNA library by modifying a GenomeWalker ligation-mediated method suitable for our purpose. 3' T overhang at the 3' end of the newly-designed GenomeWalker adaptor enhanced an adaptor-DNA fragment ligation efficiency and the long nested oligos allowed us to use next-

generation Illumina sequencing technique by the presence of Ilumina SBS sequencing primer and flow cell complimentary primer. In order to diminish unnecessary *Mu* flaking sequences transmitted from parents, we tested the efficiency of kamchatka crab duplex-specific nuclease (DSN) treatment. Because DSN normalization causes degradation of the double-stranded fraction formed by abundant PCR products, it was expected to reduce parental insertion background. However, DSN treatment was minimal in effect to reduction of parental insertion flanking sequence but induced unwanted removal of new germinal insertions. As a result, this step was skipped from the modified library generation protocol to preserve more valuable germinal insertions. Later, exceedingly fast advance of Illumina sequencing technology allowed us to have more than 100 million reads that can enough cover minimal read numbers of new germinal insertion flanking sequences despite of dominance of parental insertions.

Bioinformatic analysis make possible to distinguish germinally-transmitted *Mu* flanking sequence reads from dominating parental insertion reads. On the basis of our pilot study on 36 individual plants, we identified about 5800 unique germinal insertion candidates, which is on average 162 new germinal insertions per MTM plant. Based on this UGI number, we expect to identify about 370,000 (162 X 2,304) new *Mu* alleles in one grid (Grid M04) consisting of 2,304 MTM plants. As the maize gene space is estimated to 350–400 Mb that is about 16-17% of maize genome (Martienssen et al., 2004; Palmer et al., 2003), 370,000 insertions from Grid M04 may theoretically have 1X coverage of recovering a *Mu* insertion in any given 1 kb gene. To date, estimate of *Mu* copy number has been in a range of 50 to 100 copies (Alleman and Freeling, 1986; Bennetzen, 1984). However, the estimation on *Mu* copy number can be variable depending on backgrounds into which *Mu*

active lines are outcrossed and also *Mu* elements whose TIR regions are recognized by *Mu1* hybridization, otherwise being excluded (Alleman and Freeling, 1986; Bennetzen, 1984). Taken together, the number of germinally-transposable *Mu* elements can be more than 100, consistent with our data. Due to missing UGI coordinate information on a few chromosomes in column 2, 3, and 4, the average UGI number may be re-adjusted. Still, we admit that bioinformatically-identified UGI candidates need to be validated for authenticity of UGIs. However, our goal is to provide sequence-indexed database of new *Mu* alleles for science community who is always looking for more new alleles to study gene function and genetic networks. Since 96 libraries from one grid have been sequenced, an intensive bioinformatic analysis will create sequence-indexed MTM *Mu* FST database that may contain about 370,000 coordinates of *Mu* insertion sites.

Members of the *Mutator* family have some features that are distinctive from other transposons; 1) sequences of exceptionally long TIRs become more divergent over time, 2) internal promoters can be present in both TIRs that may direct transcription in convergent orientations, and 3) internal region of *Mu* elements are often composed of diverse captured host sequences surrounded by both TIRs (Bennetzen and Springer, 1994; Lisch, 2002). Based on our analysis, 572 parental insertions were detected from the 12 libraries sequenced. If only 162 *Mu* elements are actively transmitted, the remaining 410 parental insertion sites may be occupied by immobilized *Mu* elements whose TIRs are diversified. It is likely that the excessive immovable *Mu* elements are Pack-MULEs. (Jiang et al., 2011) actually identified 276 maize Pack-MULEs via comparing high-quality genomic sequences with gene annotation and found that the average sequence identity of TIRs of Pack-MULES is 87% for maize. If the identity of the TIRs is a critical factor for transition medicated by MuDR

transposase, maize Pack-MULEs carrying diversified TIRs may not be capable of being transposed. Also, amplification of *Mu* copy number is not striking because *Mutator* elements are germinally transmitted via replicative mechanism and duplicate their copy numbers in the pregerminal cells, meiotic cells, and haploid gametophytic cells (Alleman and Freeling, 1986; Lisch et al., 1995; Walbot and Rudenko, 2002). Over times, *Mu*-active lines may have increased *Mu* copies to prosper in the host genome whereas the host genome have developed to suppress expansion of *Mu* elements, probably resulting in *Mu* copies that have been added yet immobilized. (Jiang et al., 2011) also found that the majority of Pack-MULEs are within 500 bp from 5' termini of adjacent genes, which is consistent with our data that parental insertion sites are preferentially toward 5' upstream of start codon.

Lastly, we found that frequency of *Mu* insertion sites within each chromosome is inversely related with DNA methylation level in symmetric CG and CHG contexts. Our results support that *Mu* insertion sites are found more frequently toward the arms that are almost demethylated, yet rarely detected around centromere(s) that are highly methylated, consistent with recent findings (Liu et al.,2009; Gent et al., 2012). Although we do not provide histone modification patterns genome-widely in this study, histone modifications may affect *Mu* transposition selection due to a close cross-talk between histone modification and DNA methylation.

Figure 2.1. Schematic diagram of MTM grid sequencing.

(A) DNA from column and row is digested with three methylation-sensitive and two methylation-insensitive restriction enzymes in five separate tubes and combined them into one tube. The fragmented DNA is ligated to 4 nt barcoded adaptors. Following the ligation of an adaptor, the first PCR step is carried out with an adaptor primer (AP1) along with a primer homologous to the TIR region of the Mu transposon (TIR4). Modification of the adapter ensures that only those products primed by the Mu TIR primer can be further amplified by the adapter primer. To enrich DNA flanking *Mutator* insertion sites, the secondary PCR was performed with a nested PCR primer pair incorporated with Iluuminal sequencing primer sequence and flow cell primer sequence, which are P5-SBS3-AP3 adaptor and P7-SBS8-TIR2, respectively. The purified PCR products were paired-end (PE) sequenced for the 1st end with a library-specific adaptor primer and for the 2nd end with a SBS8 Illumina sequencing primer. **(B)** Structure of the GenomeWalker adaptor. The amine group on the short strand of the adaptor limits extension of the 3' end of the adaptor-ligated DNA fragemets and suppresses formation of a panhandle structure generated by adapter end-to-end amplification. NNNN signifies four nucleotides for barcode combinations.

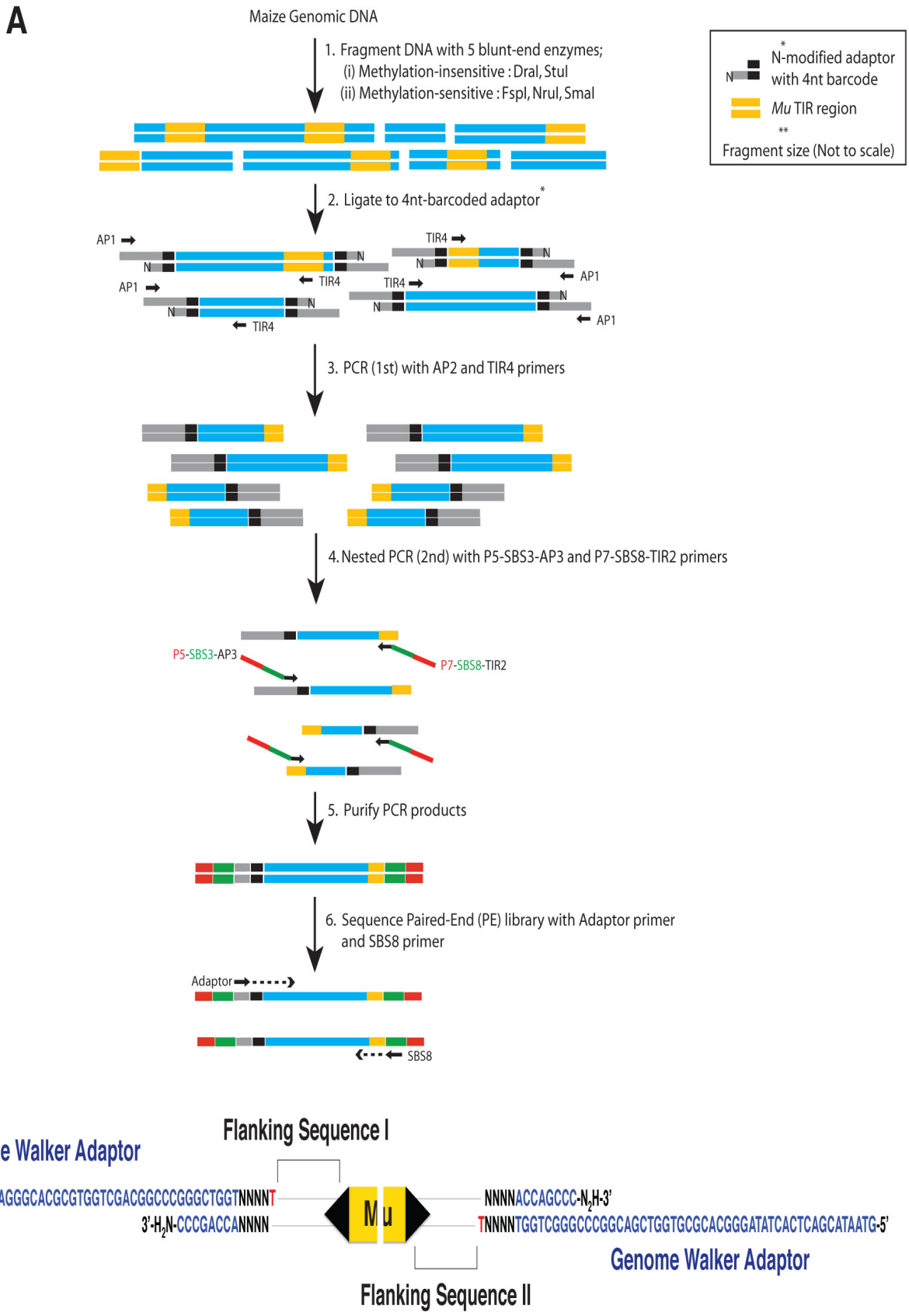
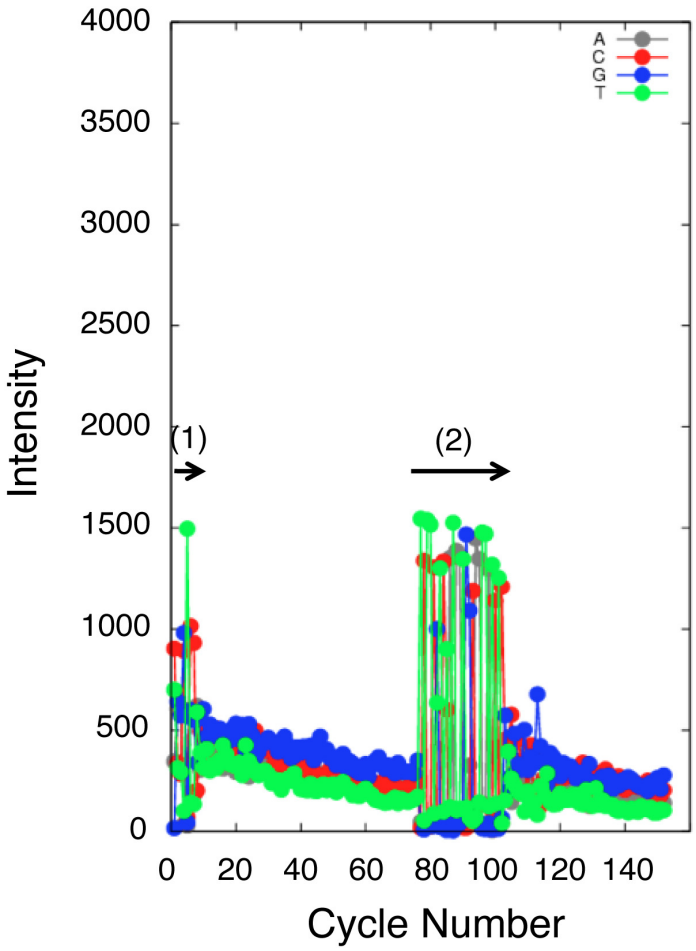

A
Maize Genomic DNA
1. Fragment DNA with 5 blunt-end enzymes;
(i) Methylation-insensitive : DraI, StuI
(ii) Methylation-sensitive : FspI, NruI, SmaI
2. Ligate to 4nt-barcoded adaptor*
3. PCR (1st) with AP2 and TIR4 primers
4. Nested PCR (2nd) with P5-SBS3-AP3 and P7-SBS8-TIR2 primers
5. Purify PCR products
6. Sequence Paired-End (PE) library with Adaptor primer and SBS8 primer
AP1
TIR4
N
P5-SBS3-AP3
P7-SBS8-TIR2
Adaptor
SBS8
*
N-modified adaptor with 4nt barcode
Mu TIR region
**
Fragment size (Not to scale)
B
Genome Walker Adaptor
Flanking Sequence I
5'-GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTNNNNT
3'-H2N-CCCGACCANNNN
Mu
NNNNACCAGCCC-N2H-3'
TNNNNTGGTCGGGCCCGGCAGCTGGTGCGCACGGGATATCACTCAGCATAATG-5'
Genome Walker Adaptor
Flanking Sequence II

Figure 2.2. Illumina Paired-End (PE) sequencing read-clusters and Mapping strategy.

(A) Illumina PE sequencing read-cluster intensity is plotted to cycle number from a single lane of GenomeWalker PCR products. The length of the 1st sequencing read is 76 nt that begins with 4 nt barcode and T adaptor overhang followed by 71 nt flanking sequence. The 2nd sequencing read starts with 26 nt *Mu* TIR sequence and the next 50 nt flanking sequence provides the exact position of *Mu* insertion sites. **(B)** Parental insertions can be extracted out by parallel comparison of pooled columns (yellow box) or rows (red box), respectively. Cross-check method (white circle) by intersecting each column and row identifies germinal Insertions in a single plant whose address is determined by column X row. 36 individual MTM lines can be pinpointed by intersecting 6 rows and 6 columns.

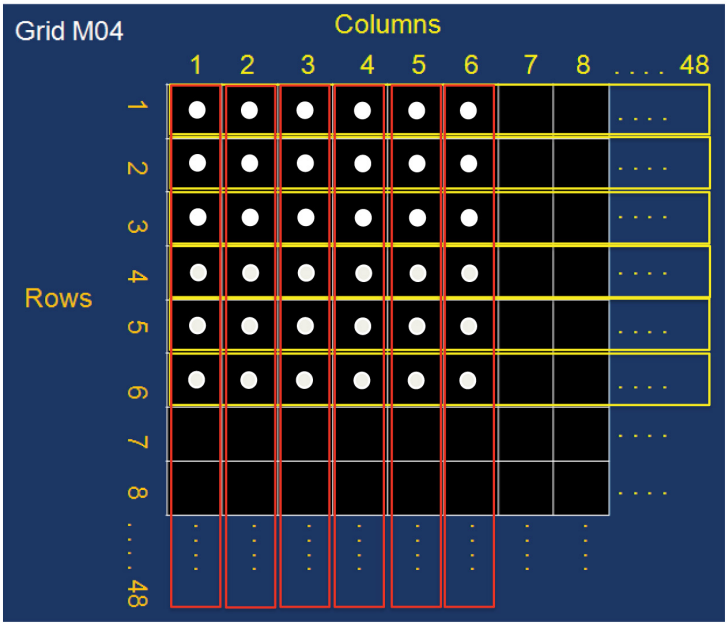
A



(1) 4 nt Barcode + T

(2) *Mu* TIR

B



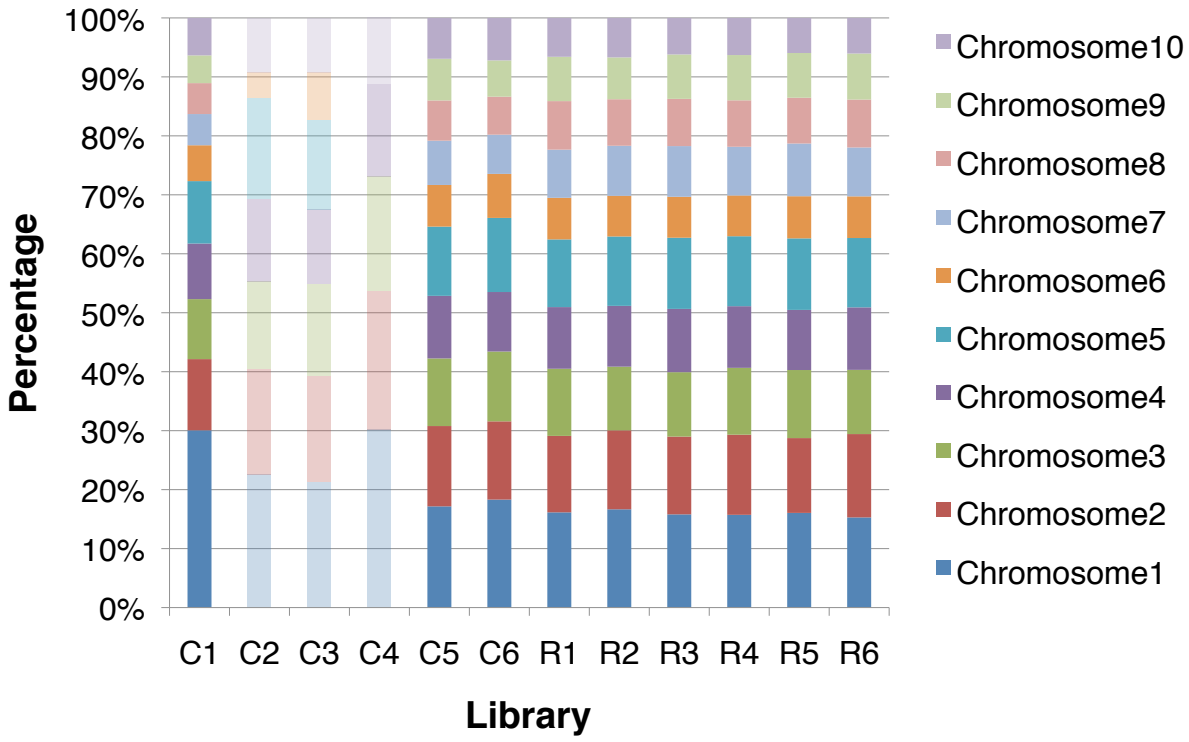


Figure 2.3. Inter-chromosomal proportion of overall MTM *Mu* target sites.

MTM *Mu* sites obtained from 12 libraries (from 6 columns and 6 rows) are found in 10 maize chromosomes in a similar proportion (Number of *Mu* insertion sites in each chromosome per total number of *Mu* insertion sites in all 10 chromosomes). The faint color bar indicates that three libraries (Column 2, 3, and 4) have insertion sites missed in a few chromosomes. For unknown reason, *Mu* insertion sites are not detected on chromosome 7 to 9 in Columns 2 and 3, even on chromosome 5 to 9 in Column 4.

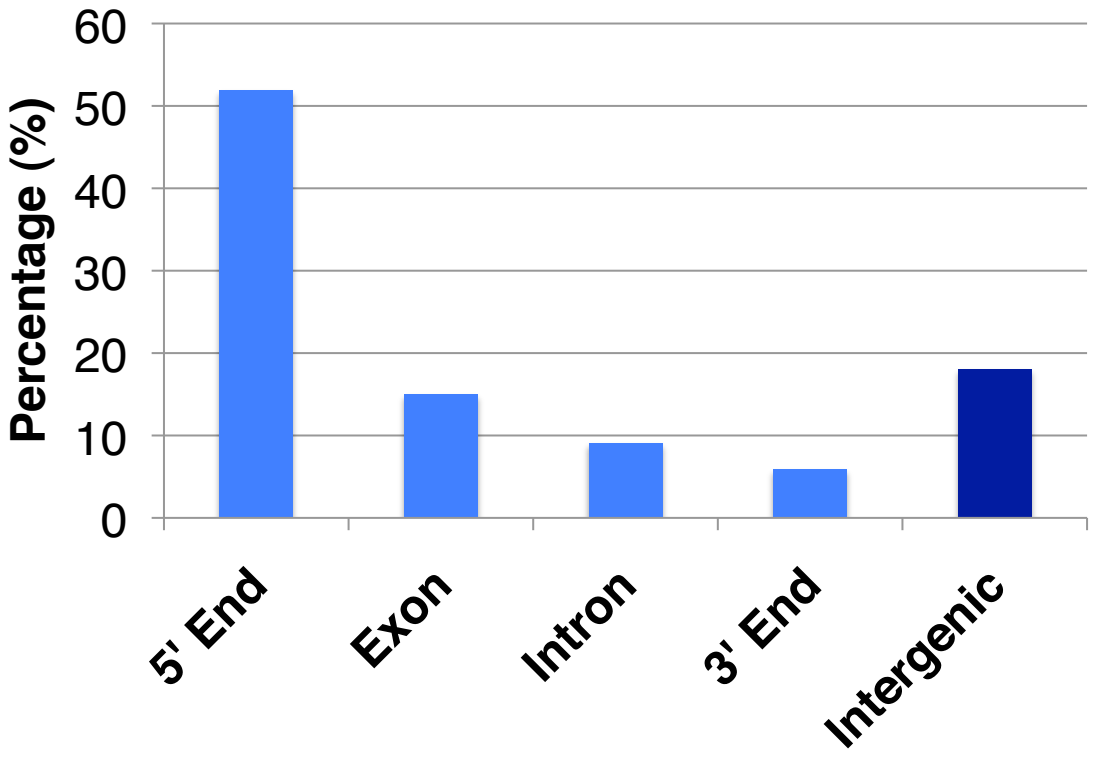


Figure 2.4. Distribution of MTM *Mu* insertions in genic regions.

The insertions were collected from 538 Parental Insertions. The proportion of MTM *Mu* target sites is a ratio of number of *Mu* insertion site in each region per total 538 parental insertions. 5' end represents 1 kb upstream from start codon, 3' end 1 kb downstream from stop codon, intergenic region further upstream from 5' end and downstream from 3' end.

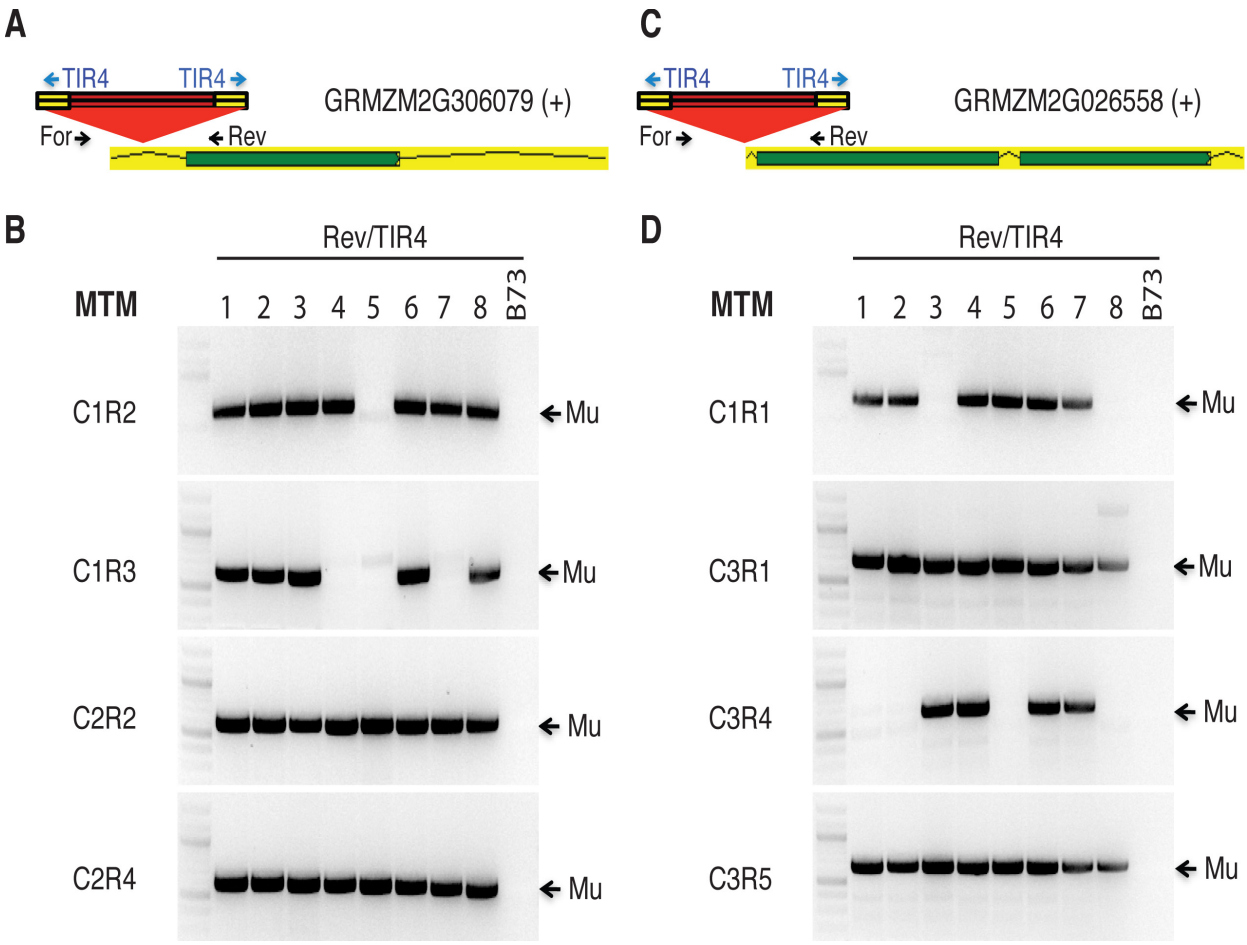


Figure 2.5. Examples of PCR validation assays for Parental Insertion flanking sequence tags (PI FSTs).

(A) and **(C)** Schematic structure of GRMZM2G306079 and GRM2G306079 with MTM *Mu* insertion site, respectively. Green box signifies exon and red triangle *Mu* element. (+) represents that the gene is transcribed in a forward orientation. Black arrow means a locus-specific primer and blue arrow MuTIR4 primer. **(B)** and **(D)** Each panel displays PCR products that were amplified using a FST locus-specific primer and MuTIR4 primer. 4 individual MTM lines were randomly selected for PI FSTs. DNA of eight F3 progenies was used for this validation assay. B73 inbred DNA template control is included in all assays.

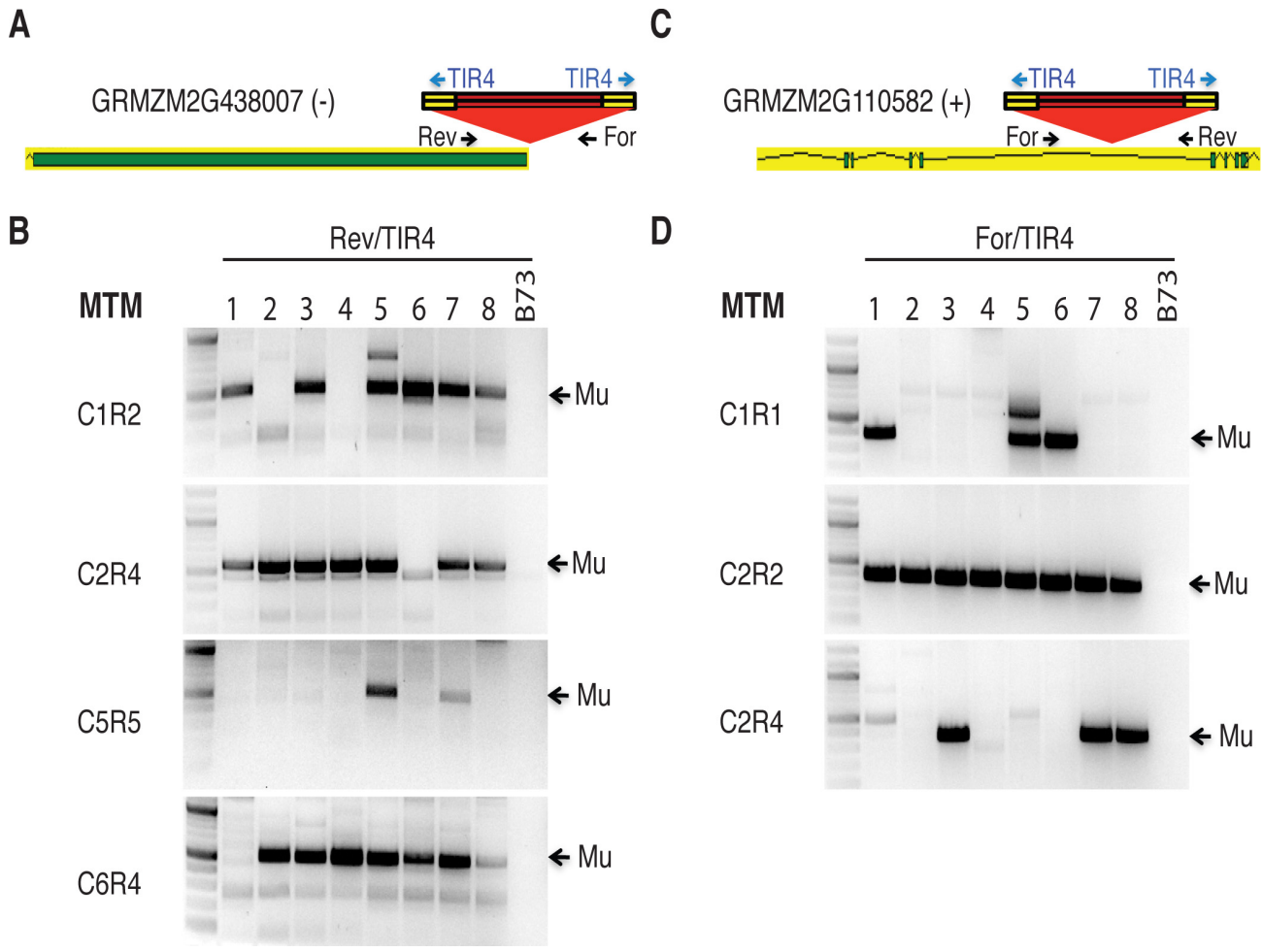


Figure 2.6. Examples of PCR validation assays for Parental Insertion flanking sequence tags (PI FSTs).

(A) and **(C)** Schematic structure of GRMZM2G438007 and GRM2G110582 with MTM *Mu* insertion site, respectively. Green box signifies exon and red triangle *Mu* element. (+) and (-) represent that the gene is transcribed in a forward and reverse orientation, respectively. Black arrow means a locus-specific primer and blue arrow MuTIR4 primer. **(B)** and **(D)** Each panel displays PCR products that were amplified using a FST locus-specific primer and MuTIR4 primer. 4 **(B)** and 3 **(D)** individual MTM lines were randomly selected for PI FSTs. DNA of eight F3 progenies was used for this validation assay. B73 inbred DNA template control is included in all assays.

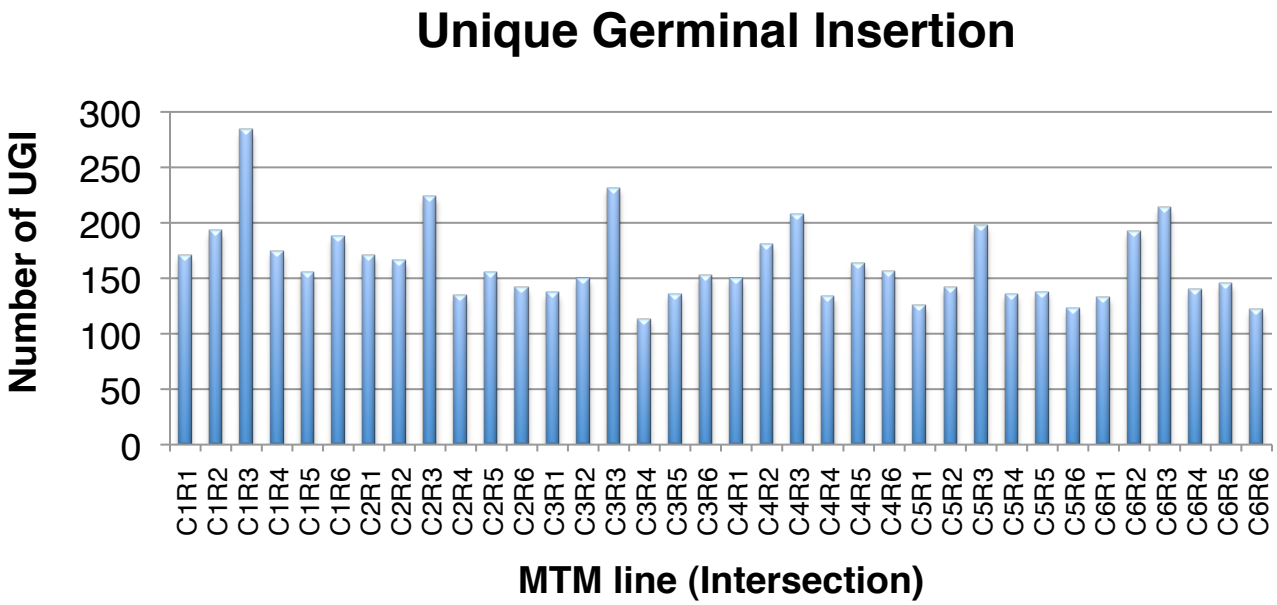


Figure 2.7. Number of Unique Germinal Insertion (UGI) from 36 individual MTM plants (intersections).

A range of 100 – 250 unique *Mu* insertion sites are identified in 36 MTM lines. Although number of UGI candidates is relatively various, they are more than 100. Cross-check (intersection) of a single column and row identifies new germinal insertions mapped to unique plants within the grid.

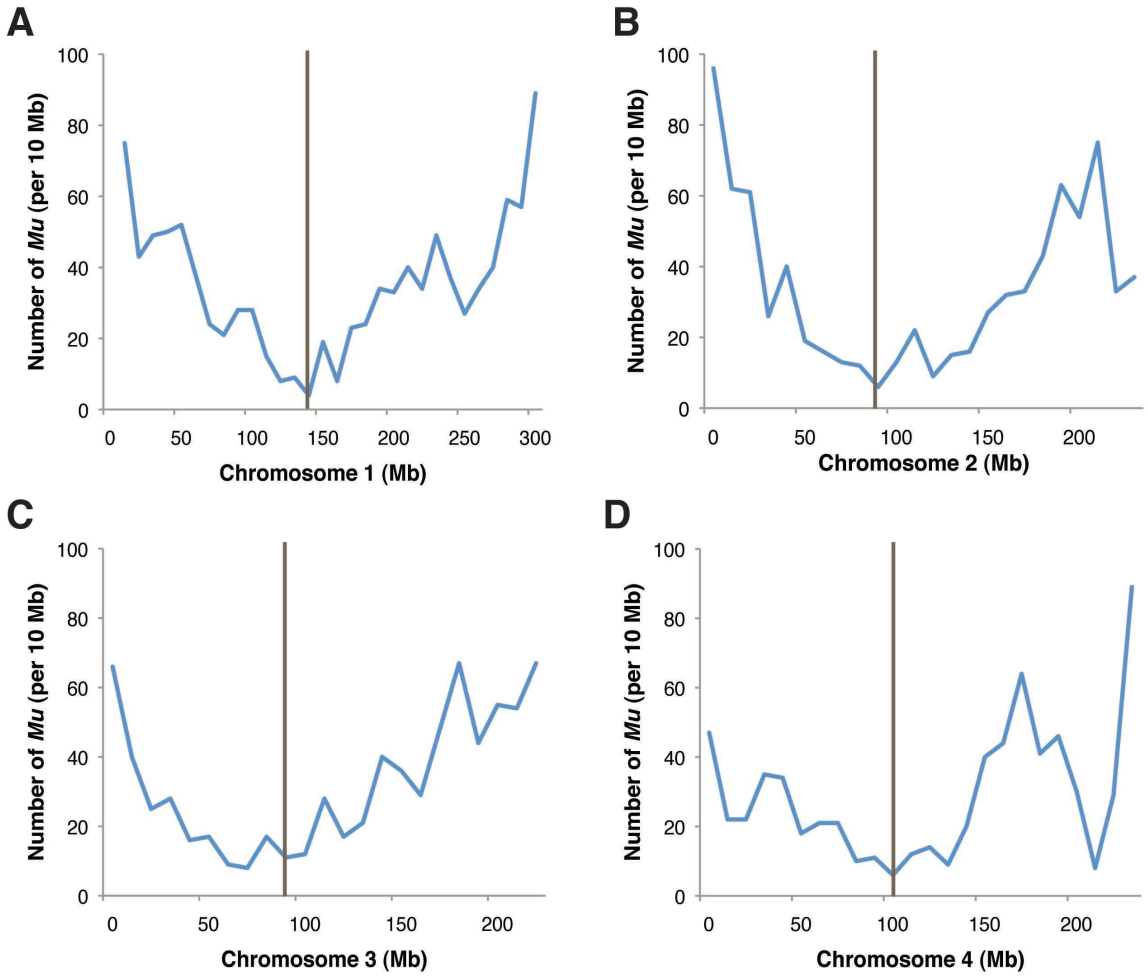


Figure 2.8. Distribution of MTM *Mu* insertions on the Maize intra-chromosomes (1 to 4).

The number of MTM *Mu* insertions per 10 Mb (Million Base) was plotted versus the coordinates (Mb) in maize chromosome 1 through 4. Since column 2, 3, and (4) missed *Mu* insertions on chromosome 7 (5) to 9, the number of all combined UGIs is minimal for plotting so that other chromosome plots are excluded. The grey line represents approximate map position of functional centromeres (Wolfgruber et al., 2009)

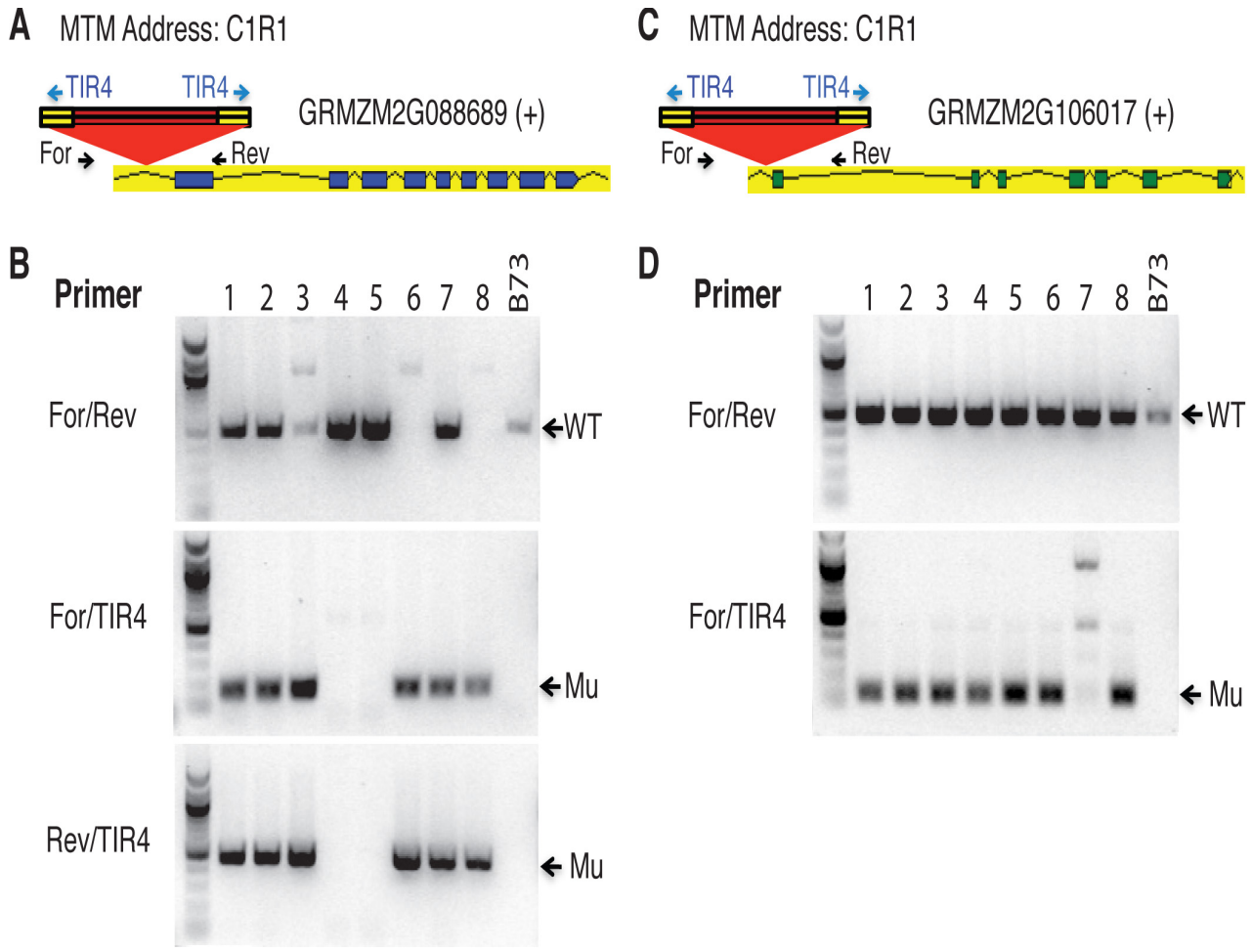


Figure 2.9. Examples of PCR validation assays for Unique Germinal Insertion flanking sequence tags (UGI FSTs).

(A) and **(C)** Schematic structure of GRMZM2G88689 and GRM2G106179 with MTM *Mu* insertion site, respectively. Two genes are UGIs in C1R1 MTM line. Blue/Green box signifies exon and red triangle *Mu* element. (+) represents that the gene is transcribed in a forward orientation. Black arrow means a locus-specific primer and blue arrow MuTIR4 primer. **(B)** and **(D)** The1st panel displays PCR products that were amplified using a FST locus-specific primer pair (Forward and Reverse primer), amplifying wild-type, the 2nd panel showing PCR products that were amplified using a locus-specific Forward primer and MuTIR4 primer, and the 3rd panel showing the amplified PCR product with a locus-specific Reverse primer and MuTIR4 primer. DNA of eight F3 progenies was used for this validation assay. B73 inbred DNA template control is included in all assays.

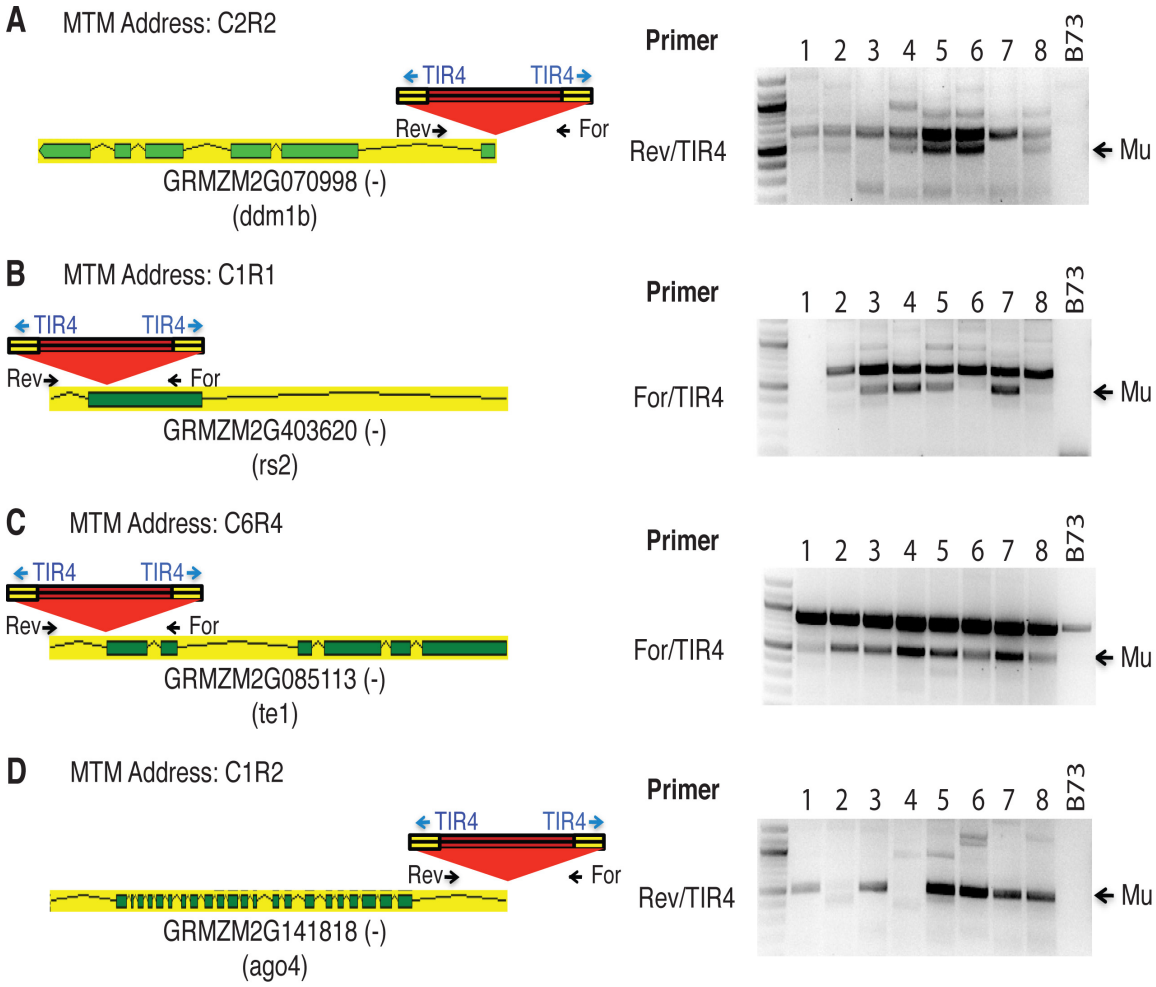


Figure 2.10. Examples of PCR validation assays for Unique Germinal Insertion flanking sequence tags (UGI FSTs) of known genes.

(A) *Mu* insertion in 5' UTR of GRMZM2G070998, which is *ddm1b* mutation. **(B)** *Mu* insertion in 1st exon of GRMZM2G403620, which is *rs2* mutation. **(C)** *Mu* insertion in 3' UTR of GRMZM2G085113, which is *te1* mutation. **(D)** *Mu* insertion in 5' UTR of GRMZM2G141818, which is *ago4* mutation. Green box signifies exon and red triangle *Mu* element. (-) represents that the gene is transcribed in a reverse orientation. Black arrow means a locus-specific primer and blue arrow MuTIR4 primer. **(A)** to **(D)** Each panel displays PCR products that were amplified using MuTIR4 primer with either a locus-specific Forward primer or a locus-specific Reverse primer. DNA of eight F3 progenies was used for this validation assay. B73 inbred DNA template control is included in all assays.

Table 2.1. Demethylation of *Mu* flanking region

Gene		Exon		MuTIR	
Context	**Methylation (%)**	**Context**	**Methylation (%)**	**Context**	**Methylation (%)**
CG	20.6	CG	9.0	CG	4.7
CHG	10.7	CHG	4.2	CHG	3.8
CHH	1.7	CHH	1.1	CHH	1.3

Table 2.2. Barcodes incorporated into GenomeWalker adaptor

GW Long PE barcode #	Barcode sequence (4 nt Barcode + 3' T overhang)
3	CTCG*T
4	TGTC*T
6	CGAC*T
9	TAGC*T
10	CAGG*T
12	ACAG*T

Table 2.3. Primers used in PCR validation assays.

Locus-specific	Primer Sequence	Figure
GRMZM2G306079 R	GCTGGTACCTCCATGGATCA	2.5B
GRMZM2G026558 R	CCGTGGTCGAAGTCATAGCTC	2.5D
GRMZM2G438007 R	AGGCACTCAGGGAGTAGTTCAG	2.6B
GRMZM2G110582 F	TGGCCTCTTCTTGTTCCTATTC	2.6D
GRMZM2G088689 F	TACCAGGATGAACCTGCTGAC	2.9B
GRMZM2G088689 R	GAGCCCAAGAACCACGAACT	2.9B
GRMZM2G106017 F	GGCGCCCTCTCTCTGTCT	2.9D
GRMZM2G106017 R	CAAACCCTTCTGAGTTTCAACC	2.9D
GRMZM2G070998 R	GAACCATGCCTTCAGCAGTT	2.10A
GRMZM2G403620 F	CCCTGGCTGTCGTCCAAC	2.10B
GRMZM2G085113 F	GCTGCCCTTCTCCTACATCTAC	2.10C
GRMZM2G141818 R	AAATGTCTCGAACCAAATCTGC	2.10D

Methods

Library Preparation

96 genomic DNA pools, 48 columns and 48 rows of Grid M04, were used for library generation (May et al., 2003). Five enzymatic reactions were set for 2 ug of 96 DNA pools each with restriction enzymes, which were DraI, SmaI, FspI, NruI, and StuI (NEB) in separate 96-well plates. The enzymatic reaction condition was set according to NEB's recommendations. The five individual reactions for 96 pools were combined into one 96-well plate and purified by Qiagen PCR purification kit. The blunt-end fragments were 3' A-tailed and ligated to 4 nt barcoded GenomeWalker adaptors (Table 2.2). To amplify *Mu* flanking sequences, 100 ng of ligated fragments were 2-step PCR enriched with Phusion High-Fidelity DNA polymerase (FINNZYMES), and adaptor-specific AP2 (5'-ACTATAGGGCACGCGTGGT-3') primer and *Mu* TIR-specific TIR4 primer (5'-GCCAWCGCCTCYATTTCGTCGAATCC-3'). The fragments were denatured at 98°C for 30 sec, followed by 35 cycles of 98°C for 10 sec, 72°C for 1 min 30 sec, and one cycle of 72°C for 5 min. To further enrich *Mu* flanking sequences selectively, the nested secondary PCR was carried out with P5-SBS3-**AP3** (5'-AAATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GTCGACGGCCCGGGCTGGT**-3') and P7-SBS8-TIR2 (5'-AAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTC CTGCTGAACCGCTCTTCCGATCT**TCTTCKTCYATAATGRCAATTATCTC-3')** using Phusion High-Fidelity DNA polymerase (FINNZYMES). The fragments were denatured at 98° for 30 sec, followed by 16 cycles of 98°C for 10 sec, 65°C for 30 sec, and

72°C for 45 sec, and one cycle of 72°C for 5 min. Amplified fragments were purified by Qiagen PCR purification kit and sequenced on an Illumina GAII platform as paired end 76 nt (PE76) reads. Six libraries with different 4 nt barcodes were loaded into a lane of flow cell. Two flow cells were used for 96 libraries. The AP3 custom primer (5'-GTCGACGGCCCGGGCTGGT-3') was used for the 1st run and Illumina SBS3 sequencing primer used for the 2nd run.

PCR Validation Assays

For PCR validation assays, genomic DNA was extracted from eight plants of each MTM F3 line and B73 inbred line. The locus-specific primers (Table 2.3) were designed with Primer3 for 20-22 base oligomers with a predicted Tm between 60-65°C. The specific primers were paired with a MuTIR specific TIR4. PCR was completed with *Ex Taq* DNA polymerase (TAKARA) and 3% DMSO added to the final reaction under optimal PCR conditions, according to TAKARA' recommendations. In general, about 100 ng of genomic DNA was denatured at 94° for 2 min, followed by 35 cycles of 94°C for 30 sec, 59°C for 30 sec, and 72°C for 40 sec, and one cycle of 72°C for 5 min. Depending on a primer specificity, annealing temperature and extension time was adjusted.

Chapter III: Association of *Ac*-derived small RNAs (sRNAs) with negative dosage effect of *Ac* element

Introduction

The kernel (seed) has played an important role in maize genetics due to the traits that can be easily screened and quantified. The aleurone, which is the outer layer of the endosperm, can display color by expressing pigmentation genes involved in the anthocyanin biogenesis (Dooner et al., 1991). DNA transposons in these genes can be used as visible markers to track transposition. The endosperm is also suitable for studying dosage because of its triploidy that enables the generation of 0 to 3 Ac copy numbers (Candela and Hake, 2008). The *Ac/Ds* elements in maize endosperm exhibit negative dosage effect that was first observed by Barbara McClintock (McClintock, 1950). "Negative dosage effect" refers to how an increase of *Ac* transposon copy number negatively affects transposition, causing less frequent and developmentally delayed excision of non-autonomous *Ds* transposon. The number of *Ac* copies within a cell is inversely correlated with the frequency and timing of transposition. In other words, the ability of *Ac* to induce *Ds* transposition is impaired more severely by three doses of *Ac* than by one dose of *Ac*. However, a positive dosage effect of *Ac* element has been found in heterologous system such as tobacco, tomato and *Arabidopsis*, as well as several *Ac* derivatives such as *Uq*, *Ac2* and *Ac-st2* in maize (Brutnell et al., 1997; Caldwell and Peterson, 1992; Dempsey, 1993; Jones et al., 1989; Swinburne et al., 1992; Yoder, 1990). Interestingly, active *wx-m7::Ac* and active *wx-m9::Ac*, which are inserted in

different locations in the *waxy* gene, respectively exhibited positive and negative dosage effects by position effects (Heinlein, 1996). These studies suggest that *Ac* dosage effect in maize is not uniform and appears to depend on *Ac* alleles and/or their inserted position.

As transposon-mediated small interfereing RNAs (siRNAs) have been characterized to be involved in transposon silencing, we hypothesize that siRNAs derived from *Ac* element may have a role in negative dosage effect of *Ac*. In general, two distinct size classes of small interfering RNA (siRNA) have been intensively studied in plants; 21-22nt siRNAs involved in post-transcriptional gene silencing (PTGS) by siRNA-guided transcript cleavage or translational repression and 24-26nt siRNAs associated with transcriptional gene silencing by RNA-dependent DNA methylation (RdDM) (Nobuta et al., 2008; Vaucheret, 2006). However, a plausible *Ac* autoregulation by siRNAs may not be enough to explain negative dosage effect because the former studies revealed that both levels of *Ac* transcript and *Ac* transposase protein increased in high dose of Ac endosperm and proposed that the negative dosage effect may be controlled by post-translational mechanism (Fusswinkel et al., 1991; Heinlein, 1996; Kunze et al., 1987). However, the mechanism regulating Ac activity is complex. For instance, Fusswinkel et al. (1991) showed a positive correlation of *Ac* TPase protein level *with Ac* copy numbers but endosperms that they used carried *wx-m7::Ac* allele that actually exhibited a positive effect according to the Heinlein's study (1996). In other words, *Ac* TPase protein level in endosperms showing negative dosage effect has not been investigated. The level of *Ac* transcript, *Ac* siRNAs, and Ac TPase protein in a dosage-dependent manner will be necessary to understand the negative dosage effect.

To study negative dosage effect in maize, we utilized a novel *Activator*, the immobilized *Ac* (*Ac-im*), which failed to make 8 bp TSDs and consequently lost 10 bp sequences at the left end, and is unable to excise itself but capable of transposing *Ds* elements germinally (Conrad and Brutnell, 2005). The *Ac-im* element is an excellent tool to understand negative dosage effect due to its crippled mobility for itself yet the *trans*-activating function for nonautonomous elements.

Results

In order to examine the negative dosage effect of *Ac-im* copy number on *Ds* excision pattern, we used lines homozygous for the *Ac-im* and *Ds* tester line *r1-sc:m3*. We self-pollinated the *Ac-im* homozygous lines to obtain 3 copies of *Ac-im* in endosperm and used the *Ds* tester lines for 0 copy of *Ac-im*. For endosperms with only 1 copy of *Ac-im*, the *Ds* tester lines carrying no *Ac-im* were fertilized with the *Ac-im* homozygous lines as a pollen donor. As shown in Figure 3.1A, when a single copy of *Ac-im* was transmitted through the male gametophyte, a very coarsely spotted aleurone phenotype was observed as previously reported (Conrad and Brutnell, 2005). Figure 3.1B also shows that *Ds* excision is delayed in timing and its frequency also is reduced in an *Ac-im* homozygous ear carrying 3 copies in endosperm tissues. This aleurone variegation was phenotypically detected at 15 days after pollination (DAP) after removing husk leaves that protect kernels from sunlight and became more obvious at 19 DAP. Taken together, we confirmed that *Ac-im* behaves in a same way as a typical *Ac* element showing the negative dosage effect.

The GenBank under accession numbers for *Ac-im*, the left-and right-end sequences flanking the *Ac-im* insertion are DQ168849, DQ068387, and DQ068388, respectively. By sequencing, we confirmed that the *Ac-im* element, identical to previously published *Ac* sequence (GenBank accession no. X05424), had lost the10 bp of *Ac* left end and have no 8 bp TSDs as well as unrelated flanking sequences neighboring each end of *Ac-im.* In lines homozygous for *Ac-im*, a *Prem-1* gypsy LTR retroelement (GenBank accession no. U03680) is located at the left end and a partial 5'UTR and exon sequence of GRMZM2G09524 at the right-end of *Ac-im*. However, according to B73 reference genome browser, the *Prem1* retroelement is 3.5 kb upstream of GRMZM2G09524 and their orientation is exclusively opposite to each other. Also, left-end flanking sequence of *Ac-im* (DQ06837) is composed of 300 bp *Prem1* retrotransposon fused to SINE element that locates at 7 kb upstream of *Prem1* in B73 reference genome. Taken together, these findings supports *Ac-im* underwent a complex integration process into chromosome 7, as suggested by Conrad and Brutnell (2005).

The *r1-sc:m3* allele has a *Ds6-like* insertion in the *r1* locus, we sequenced *r1-sc:m3* locus to determine the exact location (Alleman and Kermicle, 1993). As illustrated in Figure 3.2A and 3.2B, *Ds6-like* element is inserted in the 4th intron of *R1* gene (GRMZM5G822829) in a reverse orientation. The Sanger sequencing data for *Ac-im*, *Ds6-like,* and their flanking region were used as a reference sequence to which small RNAs were mapped with a perfect match.

Transcriptional analysis of *Ac-im* and *R1-sc* alleles

To investigate whether *Ac-im* mRNA levels corresponds to copy number, we performed qRT-PCR with endosperms carrying no *Ac-im*, a single copy of *Ac-im*, and three

copies of *Ac-im*. For this assay, primers were designed to *Ac-im* 3rd exon and 4th exon and transcript abundance was measured at two different stages of endosperm development, at 7 DAP and 9 DAP. As shown in Figure 3.3A, *Ac* transcript levels are positively related to *Ac-im* copy number at both 7 DAP and 9 DAP. As we expected, 0 dose of *Ac-im* endosperms had no detectable expression that was standardized as level '1' so that the fold change (10^3) in *Ac* expression level was easily detected in both 1 and 3 dose of *Ac-im* samples, and Figure 3.3A indicates that *Ac* is more transcribed as *Ac-im* copy number increases. Also, overall *Ac* transcript level rises as endosperms develop from 7 DAP to 9 DAP in both *Ac-im* dose samples, consistent with *Ac* transcript accumulation in the endosperm at *P-vv* from 4 DAP to 8 DAP (Brutnell et al., 1997). Therefore, *Ac* transcription itself is unable to explain negative dosage effect because more *Ac* transcripts are theoretically expected to be more translated into Ac Trasposase that are functionally responsible for *Ds* excision.

On the other hand, it has been reported in C. elegans, *Drosophila*, and mouse that naturally formed transposon-derived dsRNAs from bidirectional transcription produce small RNAs to regulate transcripts (Ghildiyal et al., 2008; Sasaki and Watanabe, 2008; Sijen and Plasterk, 2003; Watanabe et al., 2008). To test whether bidirectional transcription occurs in *Ac-im* locus, we performed qRT-PCR assays with primers specific to *Ac-im* 3' end and GRMZM2G09524, respectively. As shown in 3.3B, anti-sense *Ac-im* transcript fused with GRMZM2G09524 was detected in 1 and 3 dose *Ac-im* endosperms. Similar to sense *Ac* transcript expression pattern, this fusion transcript also accumulate more as *Ac-im* copy number increases. Because *Ac-im* is inserted into GRMZM2G09524 gene, having only one exon in a reverse orientation, it is possible that a read-through antisense *Ac* fusion transcript is generated using the promoter of GRMZM2G09524 gene. If sense and antisense *As*

transcripts are hybridized, *Ac* transcript levels may be post-transcriptionally regulated. To test whether *Ac* double-stranded RNAs (dsRNAs) are formed, RNase protection assay was carried out. However, we did not detect *Ac* dsRNA formation by far (data not shown). It is likely that dsRNAs may be quickly degraded before detection and/or target region for dsRNA detection may be wrong due to a lack of fusion transcript size information.

According to aleurone variegation phenotype (Figure 3.1), a single copy of *Ac-im* endosperms showed a sign of early and frequent *Ds* excision. To examine whether *R1-sc* allele is recovered from *r1-sc* mutation by the excision, qRT-PCR assay was performed with primers specific to exon 3 and exon 5 of *R1-sc* allele. As shown in Figure 3.3C, reverted *R1-sc* transcript was only detected in 1 dose *Ac-im* endosperms, yet not others, suggesting that pigmentation occurs within *Ds*-excised cells in the aleurone layer carrying 1 dose *Ac-im* and takes place earlier than in 3 dose *Ac-im*. This suggest that *Ds* excision out of *r1-sc:m3* allele causes the recovery of *R1* gene function and induces pigmentation in the cells where *Ds* is excised.

Small interfering RNAs (siRNAs) derived from *Ac-im* and/or *Ds*

If dsRNAs are formed from *Ac-im* and/or *Ds6-like* elements, they may be processed into different classes of small RNA. To test whether siRNAs are derived from these elements, we constructed small RNA libraries from 3 samples; 0 *Ac-im*, 1 *Ac-im*, 3 *Ac-im* endosperm RNAs at 7 DAP and mapped small RNA reads to both *Ac-im* and *Ds6-like* reference sequences including their own flanking sequence (Figure 3.4A). Both left and right borders of the *Ds6-like* element exhibit the presence of perfectly-matched 21 nt siRNA in endosperms carrying 3 *Ac-im*, more abundantly than 1 *Ac-im*, but no 21 nt siRNA in 0

Ac-im (Figure 3.4B), indicating that 21 nt siRNA matching *Ds* has a positive correlation with *Ac-im* copy number. It is possible that *Ac* transcript provides a template for imperfect dsRNA hybridization with *Ds6-like*. As we only allowed for small RNA mapping with perfect matches, 21 nt siRNA was not detected in *Ac-im* despite of only 1 nt mismatch at 3' end of both *Ac-im* and *Ds6-like* elements and 3 nt mismatches at 5' end (Figure 3.4C). If *r1-sc*:*m3* is transcribed using *R1-sc* promoter, reversely-oriented *Ds6-like* will be included in the *r1-sc* precursor mRNA (pre-mRNA) strand and anti-sense *Ds6-like* fusion transcript may be hybridized with sense *Ac* transcript to form dsRNAs. However, the 21nt siRNA mapped to 5' and 3' end is derived from the sense strand from *Ds6-like* element resembling a miRNA precursor in this respect, except that the intervening "loop" region is several kb in length (Figure 3.4C). Since the *Ds6-like* element is unable to generate a sense strand from *r1-sc:m3*, the source of this 21 nt siRNA may be other *Ds* elements that are present in unknown locations. Still, it is enigmatic how this class of *Ds*-derived siRNA is generated.

From Figure 3.5, we detected a peak of 22 nt siRNA, which is located at 5' subterminal regions about 100 bp apart from 5' end of *Ac* and/or *Ds*. As the same peak pattern was detected from *Ac-im* and *Ds* and sequence of this region is identical, it is hard to prove from which element 22 nt siRNA is derived. This peak is specific to 1 dose *Ac* endosperm, however, indicating *Ac-im* as the likely origin, and perhaps suggesting a role in promoting transposition, as the peak is absent from 0 dose and 3 dose *Ac*. In 1 dose endosperm (Fig.3.3C), *Ds6-like* element transposed from *r1-sc:m3* allele and *R1* gene expression was restored. The break at *R1* locus may be able to induce the production of DSB-induced small RNAs (diRNAs), which were discovered in *Arabidopsis* and human cells (Wei et al., 2012). Although we assumed that *Ac*- and/or *Ds*-derived 22 nt sRNA may

be involved in DNA double-strand break (DSB) repair mechanism at *R1* locus, 22 nt sRNA in a single copy of *Ac-im* endosperms appears within the *Ds* element, not around the break point as predicted by Wei et al. (2012). Thus, it is unlikely that 22 nt sRNA is mediated by DSB induction.

Finally, 24 nt sRNA are present at the 5' and 3' subterminal regions of *Ac-im* and *Ds* in all *Ac-im* dosage samples, suggesting that 24 nt sRNA are not derived from specifically *Ac-im* but from *Ds6-like* or other members of *Ds* elements. Transposition is regulated by DNA methylation of the TPase binding sites that are 5' and 3' subterminal regions (Kleckner, 1990) so that it is likely that *Ds* element may produce 24 nt sRNA to methylate subterminal regions as a defense mechanism. *Ac* dosage effect is seemingly not associated with 24 nt sRNA.

Discussion

We investigated the association of siRNAs produced by *Ac-im* and/or *Ds* elements with the most striking autoregulatory feature of Ac that is called "negative dosage effect". The reason why we began with a different angle other than the proposed post-translational model is because we found that the model to explain negative dosage effect was built on insufficient data and transposon-derived small RNAs have been proposed for transposon silencing. (Fusswinkel et al., 1991; Heinlein, 1996; Kunze et al., 1987). In this study, a higher dose of *Ac-im* produced not only more sense *Ac* transcript but also anti-sense *Ac-im* fusion transcript that could be initiated from regulatory region of the neighboring gene. This

result offers a possibility of formation of dsRNA between sense- and antisense- *Ac* transcripts, which is potentially processed into small RNAs. Also, we do not rule out the presence of *Ds*-fused mRNA with the host gene because *Ds* elements are preferentially targeted into coding regions (Vollbrecht et al., 2010). If imperfect dsRNAs are formed between sense *Ac-im* transcript and antisense *Ds* fusion transcript, they may also autoregulate *Ac* expression level post-transcriptionally. However, in RNase protection assays we did not detect double-stranded RNA formation. It is possible that hybridization efficiency of two complementary mRNA strands may be low or the hybridized dsRNAs may be already degraded even before extract RNAs from endosperm tissue. As we used a RT-PCR strategy after RNase treatment, primers designed to random position of *Ac* element might be off the size range of dsRNA.

Small RNA profiling revealed that 21 and 22 nt classes and 24 nt class of siRNAs were perfectly matched to *Ac-im* and/or *Ds6-like* elements, suggesting that siRNAs were derived from *Ac-im* or *Ds* elements. Although sRNA biogenesis pathways were not investigated in this study, one possibility is that *Ac-Ac* and/or *Ac-Ds* dsRNAs may be used as a template for siRNA generation. Also, we cannot rule out the presence of *Ds-Ds* dsRNAs formation because *Ds* elements, which are highly conserved in sequence homology, are derivative alleles of *Ac.* The perfectly matched 21 nt sRNA is not mediated from *Ds-Ds* dsRNA between *Ds6-like* and cryptic *Ds* elements due to the absence of 21 nt sRNA from 0 *Ac-im* (Figure 3.4A). Although the presence of *Ac-Ds* dsRNA may explain the production of *Ds6-like*-derived 21 nt sRNA that is sensitive to *Ac-im* dosage, the strand specificity of the siRNA suggests that other *Ds6* elements would be the source of this 21 nt siRNA.

The important feature of 1 *Ac-im* endosperm is that *Ds* transposition occurs early and frequently. At 7 DAP, *R1* gene is expressed within cells of aleurone layer, indicating that *Ds* element is already re-positioned (Figure 3.3C). Since *Ds* transposition accompanies DNA double strand break at the original site, DSB-induced sRNA (diRNA) was expected to be present in the vicinity of the break point, but kernels carrying a single copy of *Ac-im* expressed 22 nt sRNA specifically within 5' subterminal regions of *Ac-im* and/or *Ds6-like* elements as shown in Figure 3.5. 22 nt sRNA could be involved in DNA methylation of the 5' subterminal region of *Ds* transposon. Because Ac TPase recognizes the hemimethylated state of subterminal regions and transposon promoter, these 22 nt siRNA may promote transposition activity by partially methylating *Ds* regulatory regions (Schwartz, 1989). If *Ds* is inserted into new location, the 22 nt siRNA derived from *Ac*- or *Ds* may play a role in assigning *de novo* methylation site in a newly-inserted locus as well as triggering secondary RNA production (Chen et al., 2010). Consequently, 24 nt siRNA is actively involved in advancing and maintaining hypermethylated state of DNA transposons as we observed in Figure 3.6. Once the regulatory regions are hypermethylated, 24 nt siRNA overall control methylation status in *Ac-im* and *Ds* elements. As a class of 22 nt siRNA is known to play a role in heterochromatin formation as heterochromatic small RNA (Ghildiyal et al., 2008; Nobuta et al., 2008), we can not rule out a possibility of 22 nt siRNA as DNA methylation trigger.

Despite the small RNA profiling data, we do not find a direct linkage of small RNA production with negative dosage effect. Consequently, further investigation is required. Most importantly, Antibody against Ac TPase protein will be critical to decide whether a post-translational model explains negative dosage effect.

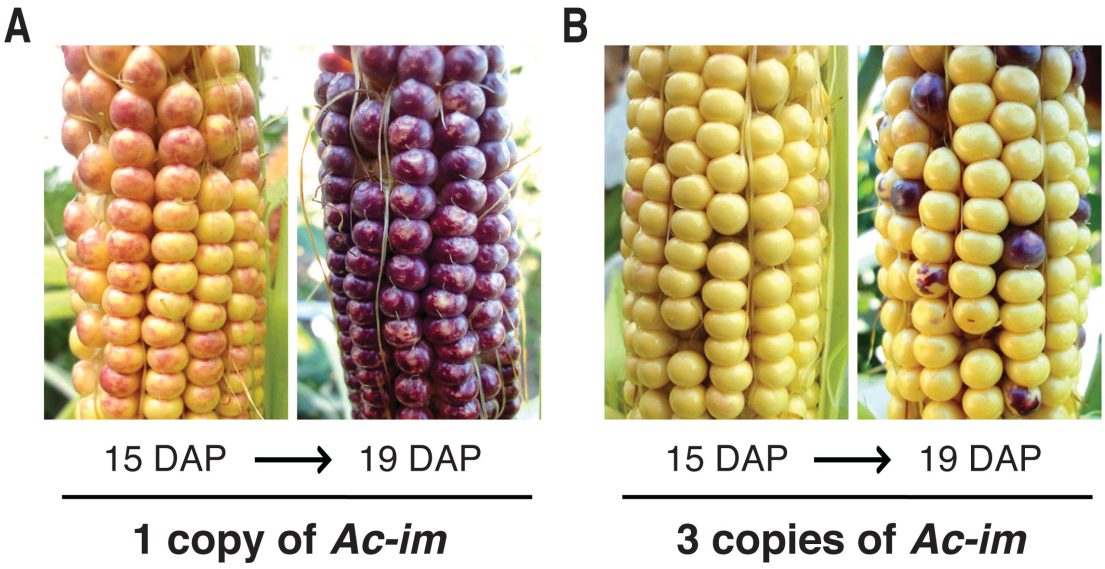


Figure 3.1. Negative dosage effect in the immobilized *Ac* (*Ac-im*).

Pigmentation patterns on aleurone layer of the triploid endosperm vary depending on copy number. Kernels having one copy of *Ac-im* shows pigmentation at 15 Days after pollination (DAP) but the colored aleurone is barely detected in three copies of *Ac-im* kernels **(B)**. At 19 DAP, pigmentation patterns become more obvious.

Figure 3.2. Structure of the *Ac-im* and *Ds6-like* elements.

(A) *Ac-im* is identical to 4656 bp *Activator* (*Ac*) except 10 bp missing at the left end. *Ds6-like* element is a derivative allele of *Ac* element, which generated a deletion between exon1 and exon, so that 10 bp deleted in *Ac-im* left end is present. **(B)** In chromosome 7, *Ac-im* neighbors Prem1 gypsy LTR retrotransopon on the left side and 1st exon of unknown gene (GRMZM2G095254) on the right side. Due to unexpected regional rearrangement, *Ac-im* lost 8 bp TSDs at both termini and 10 bp left end of *Ac* element. Red letter represents the end sequence of *Ac-im* at both ends (Conrad and Brutnell, 2005). *Ds6-like* element is positioned in the 4th intron of *R1* gene (GRMZM5G822829) in a reverse orientation. Underline signifies 8 bp TSDs. Italicized orange letter represents the sequences (8 bp TSDs and 10 bp deletion) that *Ac-im* missed but not *Ds6-like*. Arrow indicates start codon.

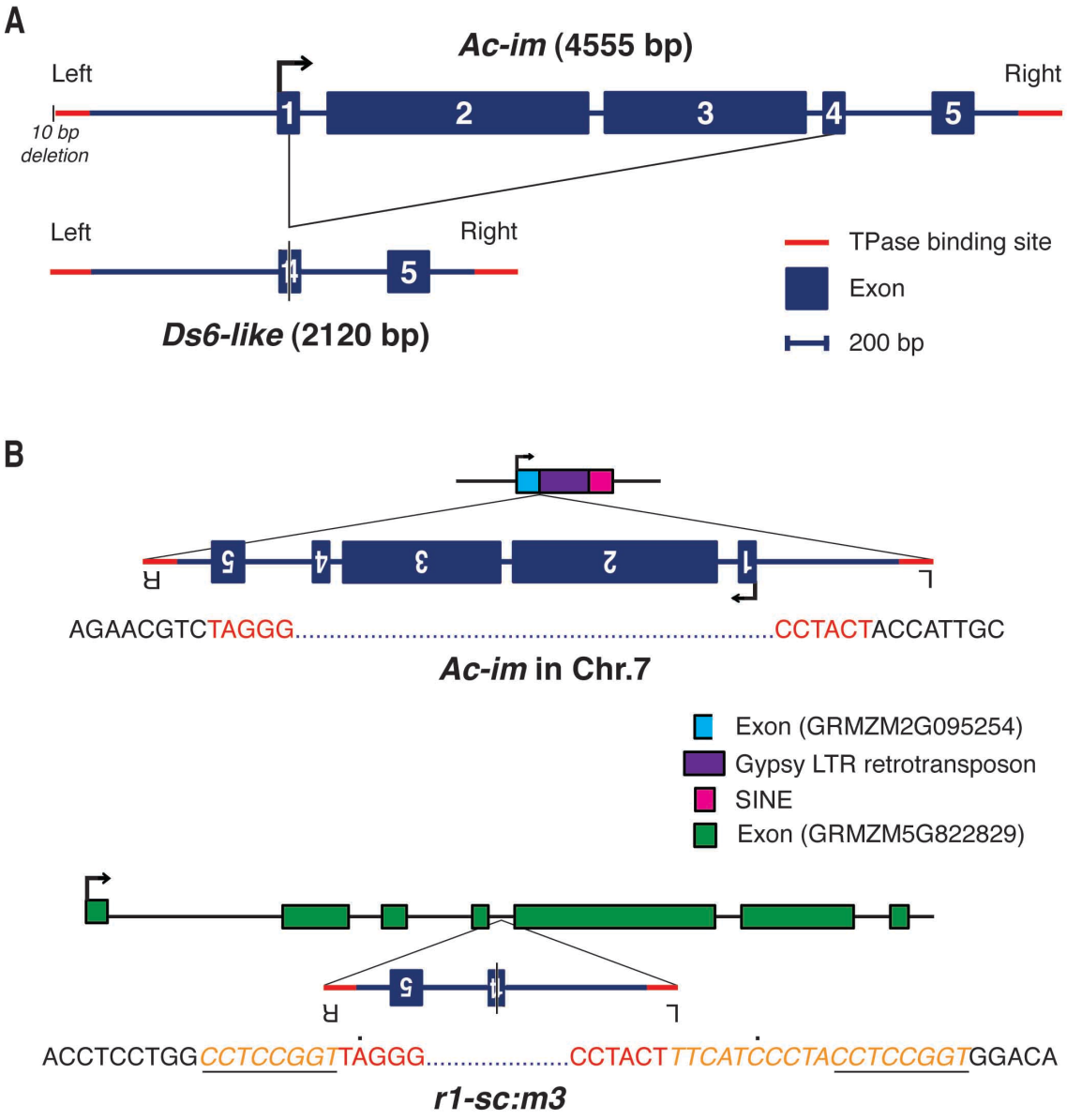

A
Ac-im (4555 bp)
Left
Right
10 bp deletion
1
2
3
4
5
Left
Right
1
5
Ds6-like (2120 bp)
TPase binding site
Exon
200 bp
B
R
L
AGAACGTCTAGGG.....CCTACTACCATTGC
Ac-im in Chr.7
Exon (GRMZM2G095254)
Gypsy LTR retrotransposon
SINE
Exon (GRMZM5G822829)
ACCTCCTGGCCTCCGGTTAGGG.....CCTACTTTCATCCCTACCTCCGGTGGACA
r1-sc:m3

Figure 3.3. Transcript abundance of *Ac-im* element and *R1-sc* reporter gene.

(A) *Ac* expression in kernels carrying zero (0 *Ac-im*), one (1 *Ac-im*), and three (3 *Ac-im*) of *Ac-im* element in 7 DAP and 9 DAP endosperm. **(B)** Anti-sense *Ac-im* fusion transcript with GRMZM2G095254 was detected in a similar pattern to sense *Ac* transcript but expression level is much lower (y-axis). Quantitative RT-PCR of two biological replicates for each sample (7 immature endosperms). Results are plotted as the ratio to the 0 *Ac-im* expression and are represented as mean +/- SEM. *Ac, Ac-im fusion, R1-sc is* normalized to *Ubiquitin* levels.

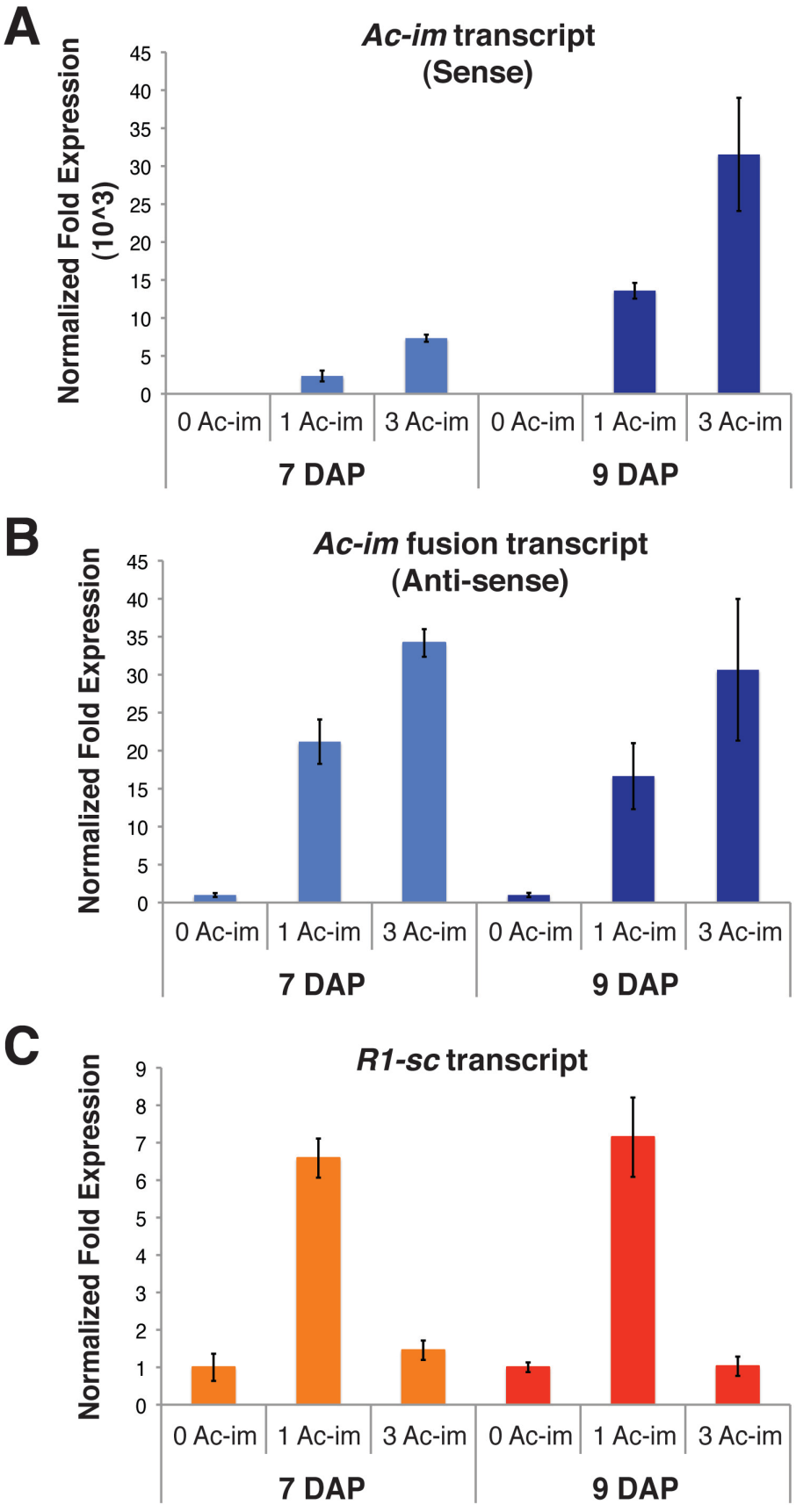

A
Ac-im transcript
(Sense)
Normalized Fold Expression
(10^3)
45
40
35
30
25
20
15
10
5
0
0 Ac-im
1 Ac-im
3 Ac-im
0 Ac-im
1 Ac-im
3 Ac-im
7 DAP
9 DAP
B
Ac-im fusion transcript
(Anti-sense)
Normalized Fold Expression
45
40
35
30
25
20
15
10
5
0
0 Ac-im
1 Ac-im
3 Ac-im
0 Ac-im
1 Ac-im
3 Ac-im
7 DAP
9 DAP
C
R1-sc transcript
Normalized Fold Expression
9
8
7
6
5
4
3
2
1
0
0 Ac-im
1 Ac-im
3 Ac-im
0 Ac-im
1 Ac-im
3 Ac-im
7 DAP
9 DAP

Figure 3.4. 21 nt small RNA abundance in *Ac-im*, *Ds6-like* elements, and their flanking regions in 7DAP endosperm.

21 nt small RNA from 0 *Ac-im*, 1 *Ac-im*, and 3 *Ac-im* endosperms at 7 DAP were mapped to *Ac-im* element with its flanking region **(A)** and *Ds6-like* element with its flanking region **(B)**, respectively. **(A)** Left-to-right end of 4555 bp *Ac-im* element corresponds to 1074 bp to 5629 bp. 1 to Small RNA reads were normalized to total reads mapped to B73 genome. Perfect match to *Ac-im* and *Ds6-like* reference sequences was counted for this analysis. Left grey line signifies left end of either *Ac-im* or *Ds6-like* element. The outside of grey line represents the flanking region. **(C)** Alignment of 21 nt sRNAs derived from the left and right end of *Ac-im* and *Ds6-like* element. Black downward arrow signifies the 1st sequence of *Ac-im* after 10 bp deletion at the left end. Purple letter represent SNPs between *Ac-im* and *Ds6-like*.

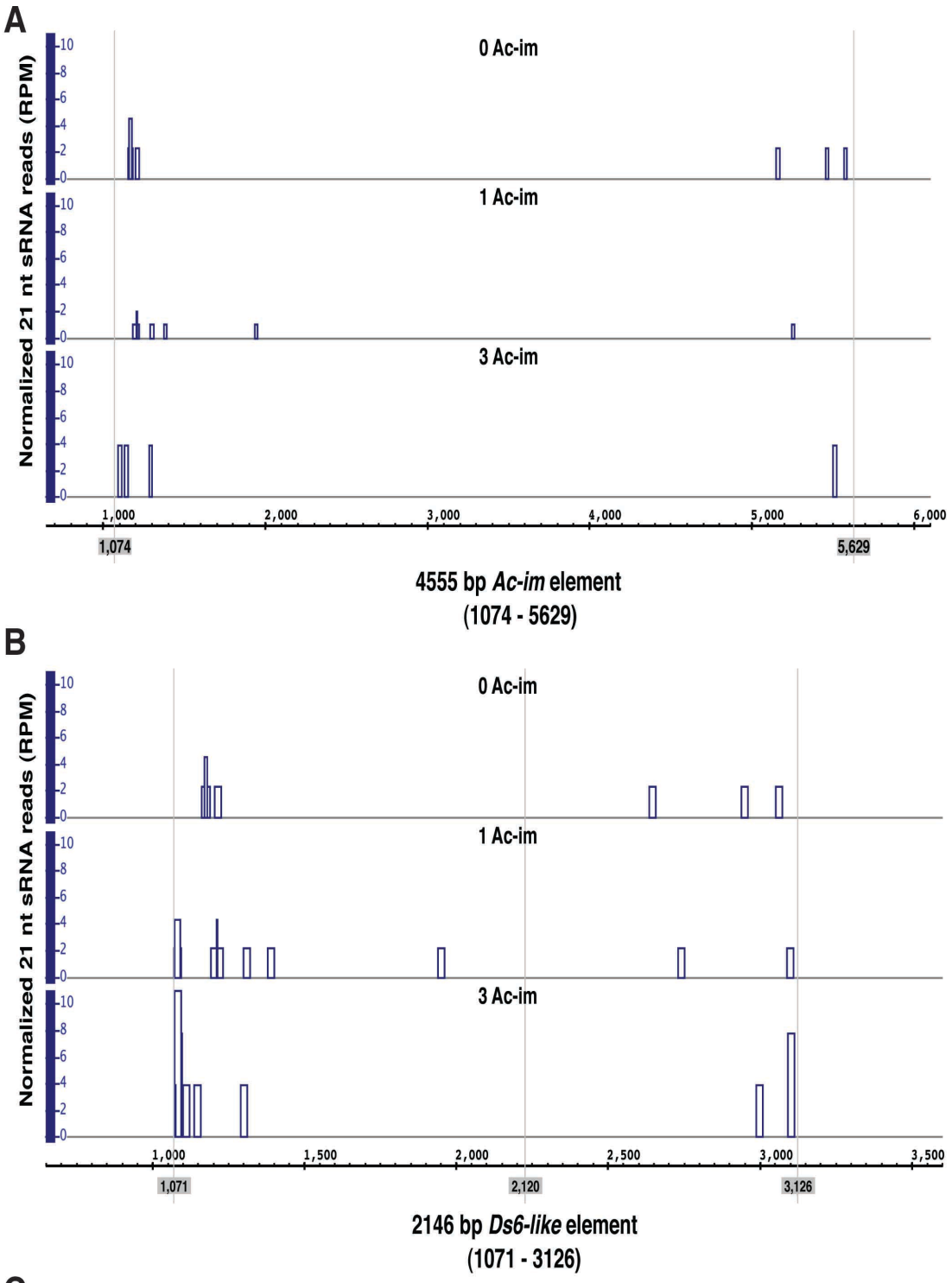

A
Normalized 21 nt sRNA reads (RPM)
0 Ac-im
1 Ac-im
3 Ac-im
1,000
2,000
3,000
4,000
5,000
6,000
1,074
5,629
4555 bp Ac-im element
(1074 - 5629)
B
Normalized 21 nt sRNA reads (RPM)
0 Ac-im
1 Ac-im
3 Ac-im
1,000
1,500
2,000
2,500
3,000
3,500
1,071
2,120
3,126
2146 bp Ds6-like element
(1071 - 3126)
C
Ac-im Left End
AATGGTAGTAGGATGGGAAAA
GATGAAAGTAGGATGGGAAAA
Ds6-like Left End
Ac-im Right End
TTACCGACCGTTTTCATCCCT
TTCCCGACCGTTTTCATCCCT
Ds6-like Right End

Figure 3.5. 22 nt small RNA abundance in *Ac-im*, *Ds6-like* elements, and their flanking regions in 7DAP endosperm.

22 nt small RNA from 0 *Ac-im*, 1 *Ac-im*, and 3 *Ac-im* endosperms at 7 DAP were mapped to *Ac-im* element with its flanking region **(A)** and *Ds6-like* element with its flanking region **(B)**, respectively. **(A)** Left-to-right end of 4555 bp *Ac-im* element corresponds to 1074 bp to 5629 bp. 1 to Small RNA reads were normalized to total reads mapped to B73 genome. Perfect match to *Ac-im* and *Ds6-like* reference sequences was counted for this analysis. Left grey line signifies left end of either *Ac-im* or *Ds6-like* element. The outside of grey line represents the flanking region.

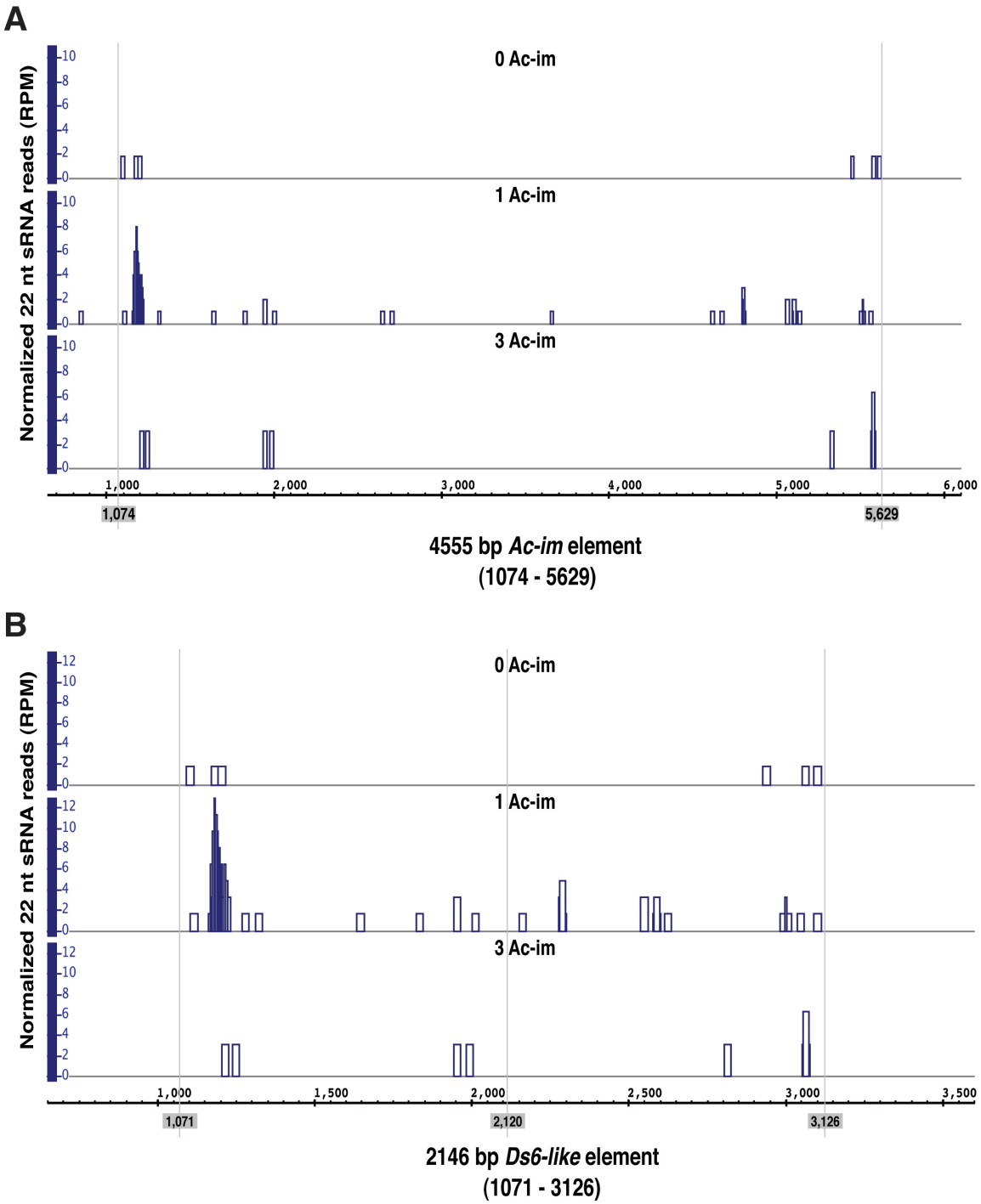
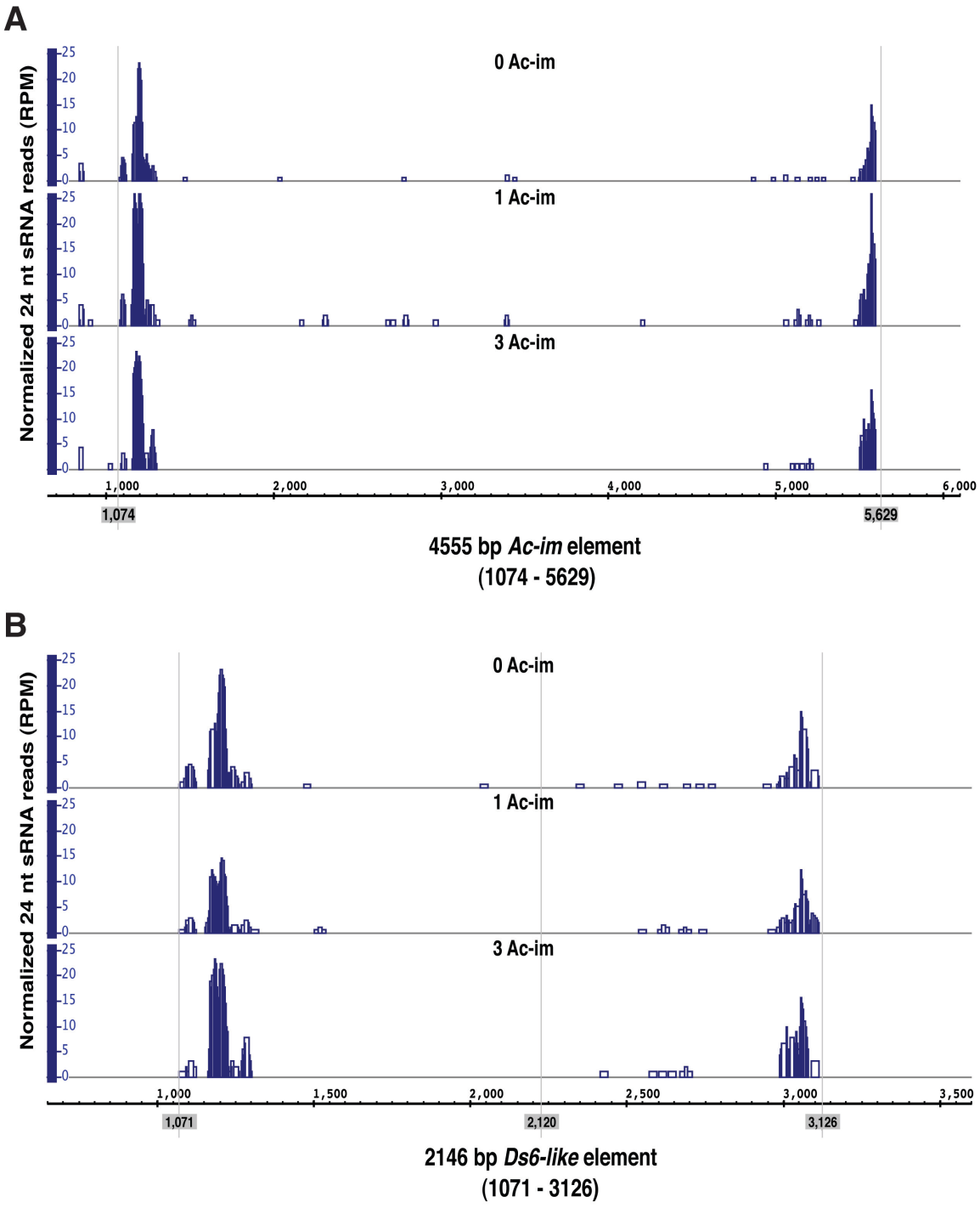

A
0 Ac-im
1 Ac-im
3 Ac-im
Normalized 22 nt sRNA reads (RPM)
1,000
2,000
3,000
4,000
5,000
6,000
1,074
5,629
4555 bp Ac-im element
(1074 - 5629)
B
0 Ac-im
1 Ac-im
3 Ac-im
Normalized 22 nt sRNA reads (RPM)
1,000
1,500
2,000
2,500
3,000
3,500
1,071
2,120
3,126
2146 bp Ds6-like element
(1071 - 3126)

Figure 3.6. 24 nt small RNA abundance in *Ac-im*, *Ds6-like* elements, and their flanking regions in 7DAP endosperm.

24 nt small RNA from 0 *Ac-im*, 1 *Ac-im*, and 3 *Ac-im* endosperms at 7 DAP were mapped to *Ac-im* element with its flanking region **(A)** and *Ds6-like* element with its flanking region **(B)**, respectively. **(A)** Left-to-right end of 4555 bp *Ac-im* element corresponds to 1074 bp to 5629 bp. 1 to Small RNA reads were normalized to total reads mapped to B73 genome. Perfect match to *Ac-im* and *Ds6-like* reference sequences was counted for this analysis. Left grey line signifies left end of either *Ac-im* or *Ds6-like* element. The outside of grey line represents the flanking region.


A
Normalized 24 nt sRNA reads (RPM)
0 Ac-im
1 Ac-im
3 Ac-im
25
20
15
10
5
0
1,000
2,000
3,000
4,000
5,000
6,000
1,074
5,629
4555 bp Ac-im element
(1074 - 5629)
B
Normalized 24 nt sRNA reads (RPM)
0 Ac-im
1 Ac-im
3 Ac-im
1,000
1,500
2,000
2,500
3,000
3,500
1,071
2,120
3,126
2146 bp Ds6-like element
(1071 - 3126)

Methods

Genetic Crosses

The *Ac-im* allele was characterized from a genetic screen for *transposed Ac* (*tr-Ac*) (Conrad and Brutnell, 2005). Lines homozygous for *Ac-im* were self-pollinated to generate endosperms carrying three copies of *Ac-im*. For a single copy of *Ac-im* in endosperm, *Ds* tester lines carrying *r1-sc:m3* were pollinated by the *Ac-im* homozygous lines. Since *Ds* tester lines did not have *Ac-im* allele, a single copy of *Ac-im* was only transmitted through the male gametophyte of lines homozygous for *Ac-im*. Although the reciprocal cross between *Ds* tester lines ad *Ac-im* homozygous lines could generate kernels carrying two copies of *Ac-im*, it was excluded in this study. As a control, *Ds t*ester lines were self-pollinated to make kernels carrying no *Ac-im* copy.

Small RNA library generation

The developing kernels were collected at 7 days after pollination (DAP) from field-grown plants; 1) lines carrying no *Ac-im*, 2) lines heterozygous for *Ac-im*, and 3) lines homozygous for *Ac-im*. From seven kernels, each embryo was carefully removed in a frozen stat and the remaining endosperms were combined for RNA extraction. To separate the starch from the nucleic acts, the SDS extraction method was first applied to maize endosperms and the TRIZOL extraction procedures were followed for the RNA enrichment (Leiva et al., 2002). MirVana miRNA isolation kit (Ambion) was used to enrich small RNAs (<200 nt) according to the manufacturer's protocol. 1 ug of small RNAs were ligated with 3' Adaptor and 5' adaptor, respectively, and followed by PCR reaction using

ScriptMiner Small RNA-seq library preparation kit (Epicentre). The PCR condition was at 95° for 30 sec, followed by 17 cycles of 94°C for 15 sec, 55°C for 5 sec, and 65°C for 5 sec. Amplified fragments of 80-100 bp were size-selected by gel extraction and sequenced on an Illumina GAII platform as single end 36 reads.

PCR Assays and Sequencing

For perfectly-matched small RNA mapping, reference sequences for *Ac-im* with its flanking region and *Ds6-like* with its flanking region were obtained by the Sanger sequencing of PCR products. 100-200 ng of genomic DNAs from lines homozygous for *Ac-im* were used to used to amplify about 6 kb PCR fragment including the left- (DQ068387) and the right-end (DQ068388) sequence spanning Ac-im (DQ168849) with Primers LC30 (5'-CCTCTCCGTTAA GTGTTTCAGG-3') and LC28 (5'- GGCAAGAA GCTTTGCTCAGA-3') (Conrad and Brutnell, 2005). The Sanger sequencing of ~6 kb PCR product confirmed that *Ac-im* insertion site and its flanking region as well as 10 bp deletion at the 5' end of *Ac-im* and the missing 8 bp TSDs at both termini. To identify insertion site of *Ds6-like* element at *R1* gene, 100-200 ng of genomic DNAs from *Ds* tester lines and *Ac-im* homozygous lines were used to amplify a junction fragment of 654 bp at the 5' left end of *Ds6-like* element with primers R1-sc Ex5 R1 (5'-AGCTC GGCTCTTCCGAGTA-3') and Ds6 R1 (5'-CTCAGTGGTTATGGATGGGAG-3'), and to amplify a 1135 bp fragment at the 3' right end with primers R1-sc Ex3 F1 (5' GACGAGCATGT CTGGCTGT-3') and Ds6 F1 (5'-CGTGCA TGGGCTATTAATTTG-3'). The PCR was performed at 94° for 2 min, followed by 35 cycles of 94°C for 30 sec, 58°C for 30 sec, and 72°C for 1 min, and one cycle of 72°C for 5 min.

qRT-PCR

Total RNA was extracted from the developing endosperms at 7 DAP and 9 DAP using the SDS-TRIZOL protocol. Total RNA was treated with DNaseI and reverse transcribed using an oligo-dT primer and SuperScript III Reverse Transcriptase (Invitrogen). qRT-PCR was performed with iQ SYBR Green Supermix (BioRad Laboratories) using two technical replicates each of two or three biological replicates. qRT-PCR primers are listed in Table 3.1. qPCR reactions were annealed at 57°C. The relative expression values for all experiments were calculated based on the expression of the experimentally validated control gene *Ubiquitin* as previously described (Satoh-Nagasawa et al., 2006). qPCR was performed on a CFX96 thermocycler and the results analyzed on the CFX Manager Software package (BioRad Laboratories). Relative expression was calculated using the 'delta-delta method' formula $2^{-[\Delta CP\ sample - \Delta CP\ control]}$, where 2 represents perfect PCR efficiency.

Chapter IV: Rearrangement at *Tunicate1* locus by *Mu*-like element

This chapter was originally published as Han et al., Plant Cell 24, 2733-2744, with modification.

Introduction

Modern cultivated corn (*Zea mays* L.), unlike other grasses in the Poaceae, has a unique feature of severe reduction in glume size that results in almost naked grains. In contrast, pod corn kernels are entirely enclosed in long glumes, resembling most grasses, and for this reason, pod corn was widely regarded as a primitive form of cultivated maize (Mangelsdorf, 1947). Pod corn was first described in the nineteenth century, as an ancient variety of maize (*Zea mays* L.) that had been preserved by pre-Columbian cultures in Brazil (Saint-Hilaire, 1829), Mexico, and Peru. Later, prehistoric cobs excavated from San Marcos Cave, dated 5,200-3,400 B.C., were found to have relatively long glumes and were thought to possibly be a weak form of pod corn (Mangelsdorf et al., 1964). Genetic mapping revealed that pod corn was attributable to a single dominant gene, *Tunicate1 (Tu1)* (Mangelsdorf and Galinat, 1964). *Tu1* behaved as a compound locus, such that rare recombinants gave rise to weak (*Tu1-l*) and intermediate (*Tu1-md*) alleles (Mangelsdorf and Galinat, 1964). A third derivative, *Tu1-d,* was phenotypically similar to *Tu1-l*, which led to the idea that weak alleles had one or two recombinable components while *Tu1* had three

components: in this way doses could vary from one to six, and recombination could first dissect and then regenerate the full *Tu1* phenotype (Langdale et al., 1994; Mangelsdorf and Galinat, 1964). Since derivatives were rare, the duplication of a recombinable component was considered as an explanation for the compound *Tu1* (Mangelsdorf and Galinat, 1964). The role of *Tu1* in the origin of corn was fraught with controversy (Mangelsdorf, 1947, 1984).

The *Tu1* phenotype is pleiotropic (Eyster, 1921; Langdale et al., 1994; Nickerson and Dale, 1955) and results in the conspicuous elongation of outer glumes, as well as sex reversal in the tassel and branching in the ear (Figure 1). *Tu1* genetically interacts with various morphological mutants that relate to juvenile-to-adult transition, such as *Corngrass1* (*Cg1*), *Teopod1* (*Tp1*), and *Teopod2* (*Tp2*); branching, such as *ramosa1* (*ra1*); and sex determination, such as the *tassel seed* mutants *ts1*, *ts2*, *ts4*, *Ts5*, *Ts6* (Collins, 1917; Langdale et al., 1994). Based on these genetic studies, *Tu1* has been proposed to play roles in phase transition, branch meristem formation, spikelet initiation, and sex determination. Despite robust genetic and morphological studies, the molecular function of *Tu1* in a diverse range of floral developmental processes is largely unknown, as is the nature of genetic modification that leads to the *Tu1* phenotype.

MADS box genes in plants are well known for regulating floral organ identity as well as developmental phase transition (Hartmann et al., 2000; Ng and Yanofsky, 2001; Whipple et al., 2004; Yu et al., 2002). A MADS box gene in maize, *Zea mays* MADS19 (*Zmm19*), lies in the same genetic interval as *Tu1* (He et al., 2004; Munster et al., 2002). Furthermore, ectopic expression of *Zmm19* results in leaf-like sepals in Arabidopsis, resembling the elongated glumes in *Tu1* mutants (He et al., 2004; Wingen et al., 2012).

Zmm19 is misexpressed in the inflorescence of *Tu1* mutants, and two duplicated copies of the gene were found in *Tu1*, but only one copy in *Tu1-l, Tu1-d* and *Tu1-md* (He et al., 2004; Wingen et al., 2012). Despite these promising indications, genome sequencing has revealed several other candidate genes in the same interval (Schnable et al., 2009a). For example, *miR172c* is found in this same genetic interval on chromosome 4 (Wei et al., 2009) and double mutants between *Tu1* and *tasselseed4* that encodes *miR172e* (Chuck et al., 2007b), are strongly synergistic (Langdale et al., 1994) with respect to both branching and sex determination. Also, *Tu1* has a synergistic interaction with *Cg1*, which encodes two *miR156* genes, *miR156b* and *c* (Chuck et al., 2007a; Langdale et al., 1994). *miR156* shows antagonistic interaction with *miR172*, which, similar to MADS box genes, has regulatory roles in the transition from vegetative to reproductive growth and in specifying floral organ identity by targeting genes involved in floral determinacy (Aukerman and Sakai, 2003; Chen, 2004; Chuck et al., 2007a; Zhu and Helliwell, 2011). As the extent of the duplication in *Tu1* was not known, the possibility that one of these other candidates was responsible for the phenotype could not be eliminated.

Here, we describe phenotypic characteristics of *Tu1* and demonstrate that the *Tu1* locus is interrupted by a chromosomal rearrangement in the 5' regulatory region of *Zmm19*. Our results indicate that the duplicate copies of *Zmm19* in *Tu1* are approximately 30 kb apart, and that they can be recombined to result in a single copy with reduced phenotypic effects. At least one other gene is included in the duplication, but *Tu1* phenotypes are reconstituted in *Zmm19* transgenic plants, when the rearranged locus is used to drive expression. TU1 protein fusions show a discrete expression pattern comparable to that of *ramosa* genes during an early stage of inflorescence development that reflects fusion of the

promoter via a large chromosomal inversion to a gene expressed in the inflorescence. Based on its expression pattern and on these dose-dependent phenotypes, we suggest that *Tu1* is involved in inflorescence architecture by promoting indeterminate cell fate.

Results

Tu1 mutant phenotypes are dosage-dependent

Tu1 was introgressed into the B73 background to examine its phenotype. Normal monoecious maize bears a distinct terminal male inflorescence, the tassel (Figures 4.1A and 4.1E), and a lateral female inflorescence, the ear (Figures 4.1D and 4.1H). Since *Tu1* is a dominant mutation, plants heterozygous for *Tu1* displayed pleiotropic defects in reproductive development, and plants homozygous for *Tu1* became even more severe in a dosage-dependent manner. Both tassels and ears of *Tu1* mutants are associated with a spikelet formation defect in which outer glumes enclosing inner whorls are highly elongated in comparison to the wild type (Figures 4.1A to 4.1I). In heterozygous *Tu1*, elongated glumes were conspicuous at the base of the central rachis and lateral branches, and became less prominent toward the inflorescence apex (Figures 4.1B and 4.1F), while homozygous *Tu1* tassels produced very large glumes (Figures 4.1C and 4.1G). In the wild type, bisexual floral meristems of maize convert into unisexual flowers by a process of selective abortion of pistil primodia within the tassel and stamen primodia within the ear (Le Roux and Kellogg, 1999). In heterozygous *Tu1* tassels, pistils failed to abort and male flowers were partially converted into female flowers near the base and some bore kernels after

fertilization (Figures 4.1B and 4.1F). This feminization was more prominent in homozygous *Tu1* (Figures 4.1C and 4.1G).

Unlike the naked kernels found in the wild type (Figure 4.1H), kernels in heterozygous *Tu1* were fully enclosed by glumes (Figure 4.1I). After removing leaf-like glumes, several long branches, which are normally found only in the tassel, were found in homozygous *Tu1* ears (Figure 4.1D, arrow). Using scanning electron microscopy of 2 mm developing ears, we observed the inflorescence meristem (IM), spikelet pair meristem (SPM), and spikelet meristem (SM) with bract growth in the wild type (Figure 4.1J). Figure 1L shows the irregular rows of the SM from homozygous *Tu1* inflorescences, which are very different from the regular pattern of the organized SM on the flanks of the wild-type ear tip (Figure 4.1J). When ears become ~ 4 mm long, the two SMs, generated from each SPM, convert into floral meristems (FM) that initiate floral organs, including anthers and pistil primordia (Figure 4.1M). Heterozygous *Tu1* ears mostly succeeded in FM conversion from SM, only producing an indeterminate branch rarely (Figure 4.1N, arrow). However, despite the obvious glume elongation in homozygous *Tu1*, we could observe numerous indeterminate long branches emerging as a consequence of failure in FM transition (Figure 4.1O), suggesting that the *Tu1* ear adopts features of the male inflorescence (Bortiri and Hake, 2007). These phenotypes suggest that *Tu1* mutants have defects in suppressing glume formation, and in meristem fate and sex determinacy.

Fine Mapping of *Tu1*

The *Tu1* mutation was mapped to the long arm of chromosome 4 (Mangelsdorf and Galinat, 1964). We set out to fine-map the *Tu1* locus by backcrossing *Tu1*/+ heterozygotes

to wild-type plants twice. We screened 738 F2 plants using markers from the region and *Tu1* was subsequently delimited between one cleaved amplified polymorphic sequence (CAPS) marker for GRMZM2G386088 and one Insertion/Deletion (InDel) marker for GRMZM2G081318, with only one recombinant each, covering a 2.5 Mb interval. To further narrow down the interval, we used two InDel markers, two insertion-deletion polymorphism (IDP) markers, and one simple sequence repeat (SSR) marker. Interestingly, these five genetic markers showed no recombination within a 1.8-Mb interval revealing a recombination "cold-spot" proximal to *Zmm19* that did not recombine with *Tu1* (Figure 4.2A). Sequencing of BAC clones in a 22 Mb interval surrounding *Zmm19* (Wei et al., 2009) allowed comparative DNA sequence analyses between the *Zmm19* gene copies found in *Tu1*, known as *Tu1-A* (Genbank AJ850302) and *Tu1-B* (Genbank AJ850303), and the *Zmm19* copy found in the reference sequence of the B73 genome. This analysis revealed that both *Tu1-A* and *Tu1-B* are structurally rearranged by insertion of a novel 2-kb *Mu-like* element in the 5' *cis*-regulatory region of *Zmm19,* which is fused with the 3' flanking region of an unknown gene (GRMZM2G006297) located on the other side of the 1.8 Mb interval from *Zmm19*, in the opposite orientation (Figure 4.2B).

This 1.8-Mb chromosomal inversion would be expected to inhibit recombination, and likely accounts for the cold-spot (Figure 4.2A) that prohibited previous attempts to precisely map *Zmm19* relative to *Tu1* (He et al., 2004; Munster et al., 2002; Wingen et al., 2012). The analyses also revealed that *Tu1-A* is distinguished from *Tu1-B* by the presence of a 3.5-kb insertion in the first intron, a non-LTR retrotransposon (Figure 4.2B, yellow triangle). Allele-specific PCR confirmed that the wild type has a single copy of *Zmm19* (Figure 4.2C), while *Tu1* has both *Tu1-A* and *Tu1-B* (Figures 4.2D and 4.2E). Additional

comparative sequence analyses and PCR showed that half tunicate mutants, *Tu1-l* and *Tu1-md,* have only one copy of *Tu1-B* or *Tu1-A*, respectively (Figure 4.2D). Finally, in one rare recombinant out of 738, we observed *de novo* crossover between *Tu1-A* and *Tu1-B* (*Tu1-rec*, Figure 4.2D), reconstructing the half-tunicate phenotype (Figure 4.3B) as previously reported (Mangelsdorf and Galinat, 1964). Since the primer pairs used in Figure 2E are separated by 1.8 Mb, only recombined *Tu1* alleles can be amplified but not the wild type. Long-range PCR and sequencing revealed that *Tu1-B* is positioned upstream of *Tu1-A,* and the intervening 30 kb DNA sequence includes *gypsy-like* and *copia-like* retrotransposons, *dSpm-like* and MITE (miniature inverted-repeat transposable element) DNA transposons, and another copy of GRMZM2G006297, fused with the promoter region of *Tu1-A*, indicating that GRMZM2G006297 is contained within the duplication (Figures 4.2E and 4.2F, and Figure 4.4). We conclude that inversion preceded duplication at the *Tu1* locus, and that GRMZM2G006297 and *Zmm19* are both strong candidate genes for *Tu1*, and cannot be further distinguished by recombination.

Zmm19 transgenic lines phenocopy *Tu1*

We used transgenic plants to determine whether the *Tu1* phenotype was caused by the rearrangement at *Zmm19*. We fused *Tu1-A* with yellow fluorescent protein (YFP) and *Tu1-B* with red fluorescent protein (RFP) at their C-termini (Figure 4.5A) and generated several independent transgenic maize plants with each transgene (see Methods). Each transgene was driven by its own promoter, which comprised 3 kb of upstream sequence, including the 3' flanking region of GRMZM2G006297 and the 2 kb *Mu-like* transposon. Maize plants were transformed individually with these two constructs (*Tu1-A:YFP* and *Tu1-*

B:RFP) and were backcrossed to B73. We performed quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis with RNA from immature tassel and ear. Each single transformant showed a significant increase in the *Zmm19* transcript abundance in both tissue types, but the relative expression level was about half of the transcript level of heterozygous *Tu1* (*Tu1*/+), which possesses both *Tu1-A* and *Tu1-B* (Figures 4.5B and 4.5C). Homozygous *Tu1* (*Tu1/Tu1*) further doubled *Zmm19* transcript abundance compared to heterozygous *Tu1* (*Tu1*/+) (Figures 4.6A and 4.6B). Expression of the transgene in transgenic immature tassels and ears was weaker than that observed in *Tu1*/+ heterozygotes, but was much higher than in wild-type inbreds. These data suggest that *Zmm19* ectopic expression levels in inflorescences are correlated with *Tu1* copy number.

Tu1-A:YFP and *Tu1-B:RFP* T1 transgenic plants produced elongated glumes in the tassel (Figures 4.7F and 4.7G), and were stable after backcrossing to B73 (Figures 4.7B and 4.7C). We tested the additive effect of the two transgenes by crossing *Tu1-A:YFP* and *Tu1-B:RFP* transgenic lines. Plants carrying both *Tu1-A:YFP* and *Tu1-B:RFP* had further elongated glumes (Figure 4.7D) and occasional feminization that enables spikelets to bear seed in the tassel (Figure 4.7E), resembling *Tu1*/+ heterozygotes (Figure 4.1F). Elongated glumes in plants with both transgenes led the main rachis and lateral branches to appear thicker than those in single transgenic and non-transgenic plants, suggesting that both transgenes are functionally involved in the *Tu1* tassel phenotype (Figure 4.7D and Figures 4.8A to 4.8F). Similarly, tassel phenotypes were more conspicuous when transgenic lines were combined with *Tu1-l*, suggesting that both transgenes interact with *Tu1-l* in a dosage-dependent manner (Figures 4.9A to 4.9G). We observed that *Tu1-A:YFP* and *Tu1-B:RFP* transgenic plants exhibited weak “half-tunicate” ear phenotypes, resembling those of *Tu1-l*

single copy derivatives (Figures 4.10A and 4.10B). Plants with both transgenes produced glumes that fully covered every kernel (Figure 4.10A), and were comparable to *Tu1/+* heterozygotes (Figure 4.1I). Furthermore, crosses between *Tu1-A:YFP* or *Tu1-B:RFP* and *Tu1-l* resulted in a dramatic enhancement of the half-tunicate phenotype in the ear (Figure 4.10B), indicating that the transgenes complemented half-tunicate derivatives and provided *Tu1* function (Mangelsdorf and Galinat, 1964). Thus, our data demonstrate that ectopic expression of *Zmm19* derived from the *Tu1* locus causes the dose-dependent *Tu1* phenotype.

Nuclear localization of TU1-A:YFP and TU1-B:RFP proteins

Zmm19 contains a highly conserved DNA-binding MADS box domain and is therefore expected to be nuclear-localized (Munster et al., 2002; Ng and Yanofsky, 2001). Confocal imaging revealed that TU1-A:YFP and TU1-B:RFP accumulated in nuclei of mature leaf and glume epidermis, as expected (Figures 4.11A to 4.11D). We further examined *Tu1-A:YFP* lines to investigate the expression of TU1 in different cell types. Nuclear-localization was detected not only in vegetative tissues, mature leaves and husks, but also in glumes, trichomes and floral meristems (FM) (Figure 4.11 and Figure 4.12). Co-localization of the two fusion proteins was consistent with their overlapping function (Figures 4.13A and 4.13B). Expression of the FP tagged proteins in vegetative tissues was expected based on the transcription pattern of *Zmm19* (GRMZM2G370777) in wild-type plants (Figure 4.14A). However, *Zmm19* was ectopically expressed in the early inflorescence of *Tu1* (Figure 4.5 and Figure 4.6), and the fusion proteins persisted in even later developmental stages, when tassel glumes were fully developed with trichomes (Figures 4.11B and 4.11D, and Figure 4.12B).

In the maize inflorescence, meristem determinacy is progressively restricted such that tertiary spikelet meristems arise from the secondary spikelet pair meristems, which in turn arise from the primary inflorescence meristem (Vollbrecht et al., 2005). While the spikelet pair is considered a short branch, long branches resemble the primary inflorescence and are normally found only at the base of the tassel. In homozygous *Tu1/Tu1* mutants, long branches also arose at the base of the ear (Figures 4.1D and 4.1O) resembling *ramosa* mutants in this respect (Bortiri et al., 2006; Satoh-Nagasawa et al., 2006; Vollbrecht et al., 2005). Both functional TU1 fusion proteins were expressed in a small cup-shaped subset of cells at the base of the spikelet pair meristem in young ear primordia (Figures 4.15A and 4.15B). Remarkably, this expression pattern is similar to that of the short branch determination genes, *Ramosa1 (Ra1)* and *Ra3,* in wild-type plants (Bortiri et al., 2006; Satoh-Nagasawa et al., 2006; Vollbrecht et al., 2005). Imaging of a double transgenic plant revealed that the YFP and RFP-fusion proteins were co-localized to the nuclei in this domain (Figure 4.15C). These images suggest that ectopic expression of TU1/ ZMM19 at the base of developing spikelet meristems promotes their abnormal indeterminacy.

Discussion

We have found that the pod corn mutant *Tu1* is caused by the ectopic expression of the MADS box gene *Zmm19* in the developing maize inflorescence. *Zmm19* is normally expressed in husk and leaf tissues in the wild type (Figure 4.14A), but is ectopically expressed in the inflorescence in *Tu1* due to a chromosomal rearrangement (Figure 4.6), most likely a large inversion associated with the transposition of a *Mutator-like* transposon.

This rearrangement led to a mild "half-tunicate" phenotype in which glumes extend but fail to enclose the kernel. Subsequently, duplication of the two genes at the breakpoint of this rearrangement enhanced the phenotype so that glumes completely covered the kernels. CACTA and MITE transposons were found at the break point of this duplication, suggesting that another DNA transposon would be involved in this regional duplication. *Zmm19* has previously been proposed as a strong candidate gene to account for the *Tu1* phenotype, but definitive proof was lacking due to the nature of the chromosomal rearrangement, which prohibited fine mapping. We have shown that half-tunicate phenotypes can be phenocopied by *Tu1* transgenes, and that these transgenes can interact with half-tunicate derivative alleles to reconstitute the full *Tu1*/pod corn phenotype.

The half-tunicate derivative alleles, *Tu1-l*, *Tu1-d* and *Tu1-md*, were recovered by Mangelsdorf and Galinat, following rare crossovers within the *Tu1* locus, and they reported that they could reconstitute *Tu1* by recombining *Tu1-l* and *Tu1-d*, suggesting that *Tu1* was a compound locus (Mangelsdorf and Galinat, 1964). They continued to characterize the half-tunicate phenotypes of *Tu1-l* and *Tu1-d* through repeated backcrossing to an inbred line and observed that *Tu1-d* consistently had longer glumes than *Tu1-l* in tassels and ears. Although the origin of *Tu1-md* is unclear, the phenotype of *Tu1-md* (Figure 4.3A) is more severe than that of *Tu1-l* (Figure 4.10B) and this may be the stronger half-tunicate allele reported by Mangelsdorf and Galinat (1964) (Langdale et al., 1994).

Previously, genomic cloning recovered two *Zmm19* genes in *Tu1* known as *Tu1-A* and *Tu1-B*, and one gene each in *Tu1-d*, *Tu1-l*, and *Tu1-md* (Munster et al., 2004; Wingen et al., 2012). Sequence analysis indicated that both *Tu1-l* and *Tu1-d* were analogous to *Tu1-B,* while *Tu1-md* was analogous to *Tu1-A*. We used PCR assays to confirm this organization

and to reveal that the *Tu1-A* and *Tu1-B* genes were part of a larger 30-kb tandem duplication in *Tu1* that included at least one other gene (Figures 4.2D to 4.2F, and Figure 4.4). It is plausible that the derivative alleles were caused by unequal crossover within this duplication, and indeed we recovered a new derivative allele *Tu1-rec,* among our *Tu1* mapping population at a frequency of 1 in 738, which is comparable to the 1 in 1,300 frequency reported previously (Mangelsdorf and Galinat, 1964). The phenotype of *Tu1-rec* is comparable to that of *Tu1-md* (Figure 4.3), and stronger than *Tu1-l* (Figure 4.10) consistent with *Tu1-md* and *Tu1-rec* retaining *Tu1-A* while *Tu1-l* retains *Tu1-B* (Figure 4.2D). We did not attempt to reconstitute the full tunicate phenotype from these derivative alleles (Mangelsdorf and Galinat, 1964), given that they only have one and not two components as previously proposed (Langdale et al., 1994), and that the mechanism of reconstitution therefore remains unclear.

By comparison with the finished sequence of 22 Mb surrounding *Zmm19* in the B73 inbred line, we detected a large chromosomal inversion in *Tu1* whose breakpoint lies in the promoter region of *Zmm19.* This breakpoint results in fusion with the 3' flanking region of an unknown gene, GRMZM2G006297. This unknown gene is duplicated along with *Zmm19*, and is expressed in both husk leaves and inflorescences, unlike *Zmm19* that is normally expressed in husk leaves (Figures 4.11A and 4.11B online). We considered three possibilities to explain why *Zmm19* expression is drastically altered in *Tu1*. Firstly, a novel *Mu-like* element located at the breakpoint of the inversion may enhance *Zmm19* expression. However, we did not detect transcription start sites within the *Mu-like* element by 5' RACE. Secondly, chromosomal inversion in the promoter of *Zmm19* may cause removal of enhancer-blocking insulators or silencers to upregulate *Tu1.* However, transgenic fusion

proteins that contained these sequences were expressed in a specific subset of cells, making this possibility unlikely. Lastly, *Zmm19* may adopt the expression pattern of the upstream gene GRMZM2G006297, accounting for ectopic expression in the male and female inflorescence. As shown in Figure 4.16.A, the GRMZM2G006297 gene expression is positively correlated to *Zmm19* dosage, suggesting that the expression pattern of the GRMZM2G006297 gene is shared with *Zmm19*.

In the wild type, RNA-seq expression data revealed that *Zmm19* expression decreases during early tassel development (comparing samples taken at a tassel length of 1-2 mm with those at 5-7 mm), while GRMZM2G006297 expression increases. During the developmental progression from IM to FM, GRMZM2G006297 expression is upregulated in young ear primodia where *Zmm19* is not normally expressed (A. Eveland, A. Goldshmidt, and D. Jackson, unpublished results). Thus, this upstream gene may be the cause of the cup-shaped expression pattern of TU1 fusion proteins, which typically resembles that of *ramosa* genes in young ear primodia. Interestingly, KNOTTED-1 (KN1) ChIP-sequencing data (Bolduc et al., 2012) revealed that KN1 strongly binds this 3' flanking region of GRMZM2G006297, which is fused with the *Zmm19* promoter by the *Tu1* rearrangement (Figure 4.16B). KN1 targets genes in the gibberellin (GA) and brassinosteroid (BR) pathways, which are involved in sex determination in maize (Bolduc and Hake, 2009; Bolduc et al., 2012; Hartwig et al., 2011). Thus, it is possible that the ectopic expression of *Zmm19* may be associated with the *Kn1* gene network.

SHORT VEGETATIVE PHASE (SVP) is the closest homolog of *Zmm19* in Arabidopsis, and regulates flowering repression (Hartmann et al., 2000), while the closely related *AGAMOUS-LIKE 24* (*AGL24*) regulates floral meristem identity and causes floral

reversion with bract-like sepals and no petals when overexpressed (Yu et al., 2004). In *Oryza sativa* (rice), the ectopic expression of Os-*MADS22,* the *Zmm19* ortholog, causes abnormal floral morphology, including loss of the palea from spikelets, elongated glumes, and a two-floret spikelet, a mild form of spikelet meristem indeterminacy (Sentoku et al., 2005). In *Hordeum vulgare* (barley), the *Zmm19* homolog *MADS1* (*BM1*), is expressed in vegetative tissues and repressed during floral development, and also induces floral reversion by repressing spike development when ectopically expressed (Trevaskis et al., 2007). Thus, the molecular role of *Zmm19* orthologs is at least partially conserved in other species. MADS-domain proteins bind CArG-box *cis*-regulatory elements (Pollock and Treisman, 1991; Shore and Sharrocks, 1995) and these are present in 5' upstream regions of *Ra1*, *Ra2*, *Ra3, Ts1*, *Ts2*, and *Ts4,* all of which interact with *Tu1*. It is plausible that ZMM19 may recognize these binding sites and contribute to pleiotropic alterations in inflorescence architecture that are exacerbated in double mutants. In addition to dramatically enhanced glume length, homozygous plants with four copies of the gene fusion have branch determination defects resembling *ramosa* mutants. These defects are consistent with the idea that *Ramosa1* and *Tu1* activity may be mutually repressive, via the co-incidence of the expression patterns of *Ra1, Ra3* and *Tu1*. Both *ra1* and *Tu1* mutants were first considered to be distinct subspecies, and Collins (Collins, 1917) described hybrids of *Zea ramosa* and *Zea tunicata* (*ra1/ra1 Tu1/*+) as sterile highly branched "cauliflower-like" inflorescences (Langdale et al., 1994). This "monstrous" phenotype (Collins, 1917) presumably reflects the co-expression described here.

Mangelsdorf famously proposed that a half-tunicate form of *Tu1* was present in the teosinte ancestors of maize, but was only revealed phenotypically in crosses to the Mexican

popcorn, Palomero Tolequeño (Mangelsdorf, 1974). Several inbred accessions of teosinte have recently been sequenced, as well as the Mexican popcorn (Chia et al., 2012). We searched for the junctions of the 1.8-Mb inversion in these genomic sequences and found no evidence for the existence of this gene fusion, although the sequences on either side were intact. While we cannot exclude the possibility that additional accessions of teosinte may have the rearrangement, our results are consistent with a single, late origin for *Tu1*, well after the domestication of maize.

Figure 4.1. *Tu1* mutant phenotypes.

(A) and **(E)** Mature tassel phenotype of wild-type plants. **(B)** and **(F)** Mature heterozygous *Tu1* tassel with partially elongated and feminized florets. Feminized florets with silk produced kernels via successful fertilization **(F)**. **(C)** and **(G)** Mature homozygous *Tu1* tassel with fully elongated and feminized florets. Single silk strands emerged from the feminized spikelets in *Tu1* homozygous tassels. **(D)** Unfertilized ears of wild-type (left), heterozygous *Tu1* (middle), and homozygous *Tu1* (right). Heterozygous *Tu1* ear displayed elongated glumes in comparison to the wild-type. Further elongated glumes and abnormal branches (arrow) were observed in homozygous *Tu1* ear. **(H)** Mature kernels were naked in the wild-type ear. **(I)** The presence of elongated glumes made kernels invisible in the mature heterozygous *Tu1* ear. **(J)** to **(L)** Scanning electron micrographs (SEM) of 2-mm-long ear development of the wild type **(J)**, heterozygous *Tu1* mutant **(K)**, and homozygous *Tu1* mutant **(L)** with irregular seed rows. **(M)** to **(O)**, SEMs of 4-mm-long ear development. Spikelet meristems (SM) in wild-type converted into a pair of floral meristems (FM) surrounded by glumes **(M)**. Heterozygous *Tu1* mutants also produced FMs normally except a few cases of an indeterminate branch (arrow) **(N)**. However, homozygous *Tu1* mutants showed failure of FM conversion and instead formed indeterminate branches with largely elongated glumes at the base of the ear **(O)**. Scale bars, 0.2 mm **(J)** to **(O)**. IM, inflorescence meristem; SM, spikelet meristem; SPM, spikelet pair meristem.

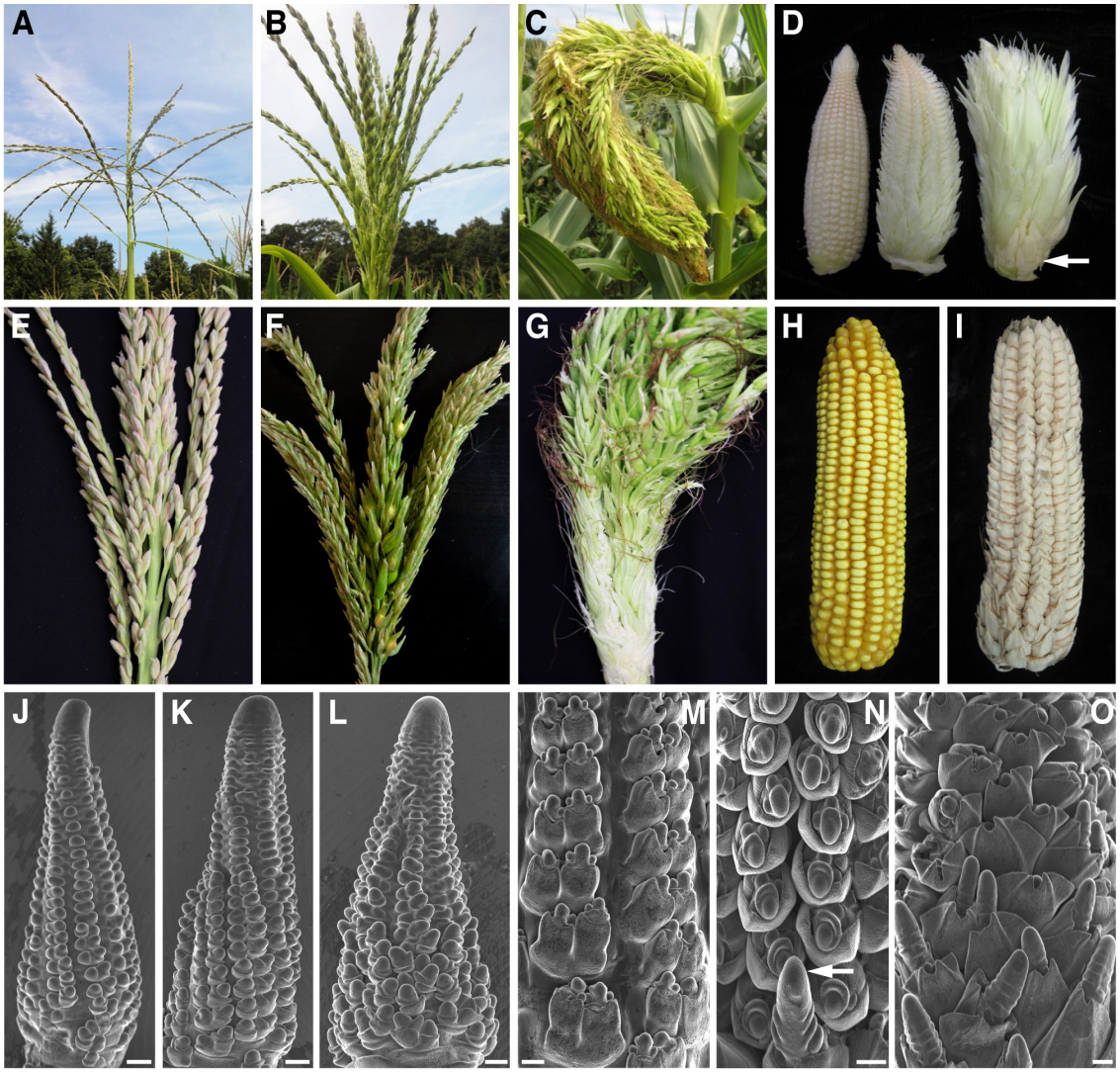
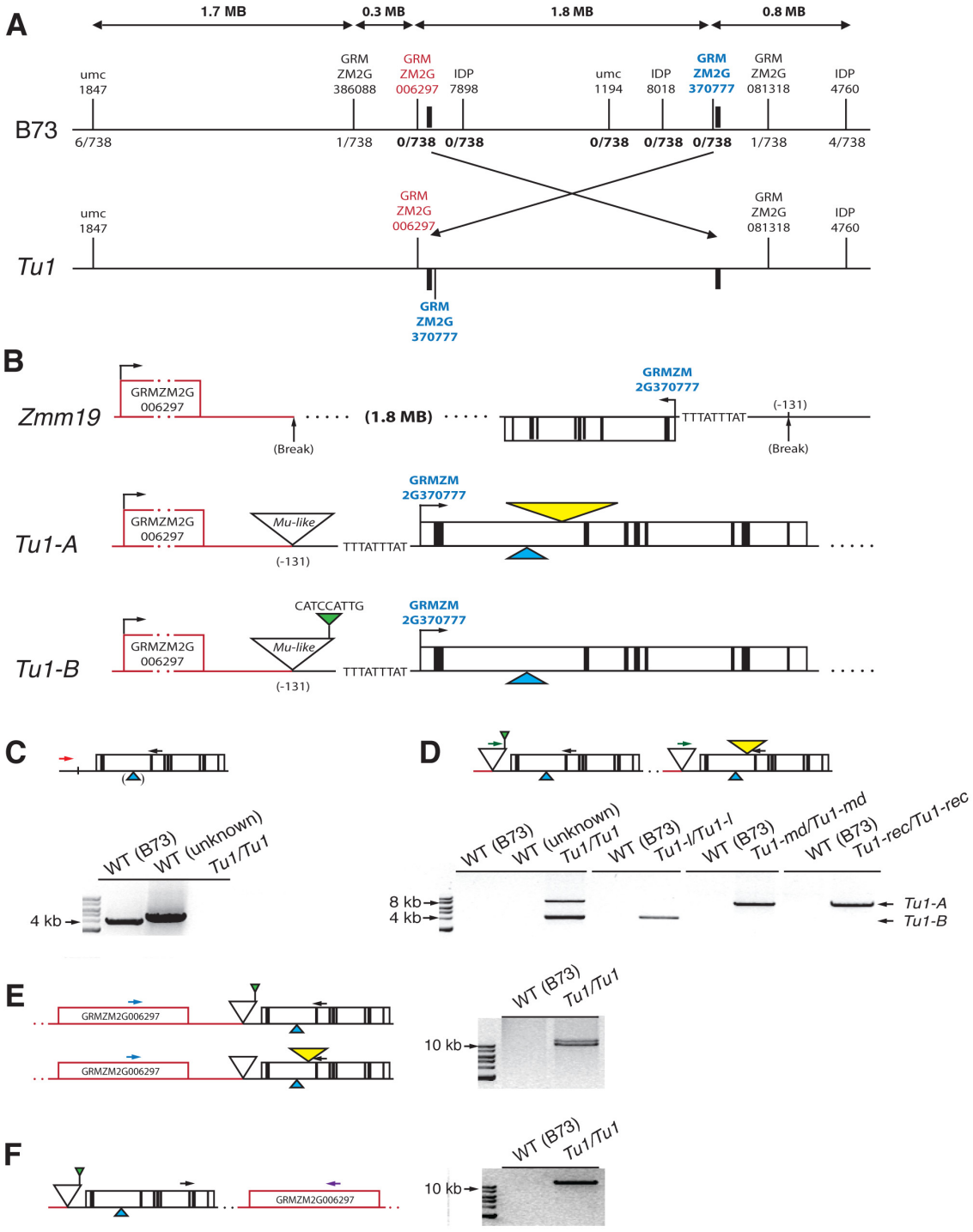

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O

Figure 4.2. Rearrangement of *Tu1*.

(A) Genetic map of the *Tu1* locus with genetic markers. Recombination frequency is indicated by number of recombinants over total population size, with a recombination cold-spot proximal to *Tu1*. **(B)** Schematic representation of *Tu1-A* and *Tu1-B* loci that are distinct from wild-type *Zmm19*, which is GRMZM2G370777. An unknown gene (GRMZM2G006297, red line), is located 1.8 Mb from *Zmm19* in the wild type, but adjacent to both *Tu1-A* and *Tu1-B* in *Tu1*. A 3.5-kb insertion (yellow triangle) specific to *Tu1-A,* a 9 bp insertion (green triangle) specific to *Tu1-B,* and a novel *Mutator-like* insertion found in both *Tu1-A* and *Tu1-B* were absent from B73. *Zmm19* is represented by a box and internal lines denote exons. **(C)** Schematic representation of wild-type *Zmm19* with 293bp insertion (blue triangle) in 1st intron, which is absent in B73 but present in unknown wild-type *tu1* background. Allele-specific PCR showing that 5' upstream region of wild-type *Zmm19* is absent from *Tu1*. The position of primers is depicted as arrows. Red arrow specific to 5' upstream region of wild-type *Zmm19*. **(D)** PCR verification of *Tu1* locus. While duplicated *Tu1-A* and *Tu1-B* loci are found in *Tu1*, half-tunicate mutants, which are *Tu1-l* and *Tu1-md*, possess one copy of duplicated *Tu1* loci, respectively. New *Tu1* derivative allele, *Tu1-rec*, was detected in our mapping population (1 out of 738) and bears only the *Tu1-A*, but not *Tu1-B*. The position of primers is depicted as arrows. Green arrow specific to *Mutator-like* insertion, which is present only in *Tu1* alleles. **(E)** Both duplicated *Tu1-A* and *Tu1-B* loci have an unknown gene (GRMZM2G006297) as their 5' upstream gene. Two amplicons are the product of two recombined genes between *Zmm19* and the unknown gene. Blue arrow specific to GRMZM2G006297. While forward primers were uniquely designed for allele-specific PCRs, the same reverse primer (black arrow) was used as in **(C)**. **(F)** Duplicated *Tu1-A* and *Tu1-B* loci neighbors each other. GRMZM2G006297 is located at the 3' downstream region of *Tu1-B* and at the 5' upstream region of *Tu1-A* (Figure 4.2E and Figure 4.4). Purple arrow specific to GRMZM2G006297. Gel images were inverted for better contrast and NEB 1kb DNA ladder was used **(C)** to **(F)**.


A
1.7 MB
0.3 MB
1.8 MB
0.8 MB
umc 1847
GRM ZM2G 386088
GRM ZM2G 006297
IDP 7898
umc 1194
IDP 8018
GRM ZM2G 370777
GRM ZM2G 081318
IDP 4760
B73
6/738
1/738
0/738 0/738
0/738 0/738 0/738
1/738
4/738
Tu1
GRM ZM2G 370777
B
Zmm19
GRMZM2G 006297
(Break)
(1.8 MB)
GRMZM 2G370777
TTTATTTAT
(-131)
Tu1-A
Mu-like
Tu1-B
CATCCATTG
C
WT (B73)
WT (unknown)
Tu1/Tu1
4 kb
D
Tu1-l/Tu1-l
Tu1-md/Tu1-md
Tu1-rec/Tu1-rec
8 kb
Tu1-A
Tu1-B
E
GRMZM2G006297
10 kb
F

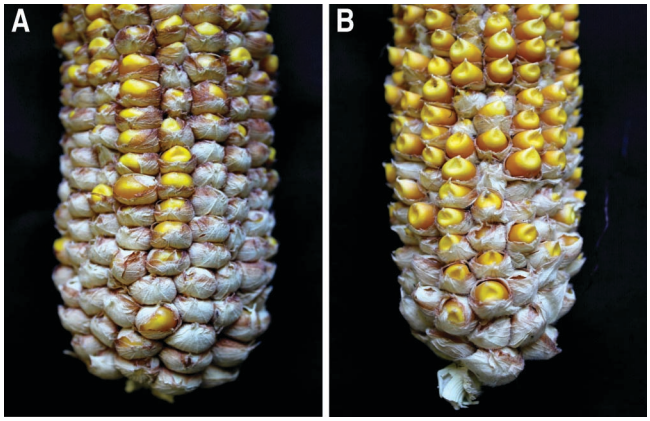


Figure 4.3. Half-tunicate phenotype of single copy *Tu1-md* and *Tu1-rec* heterozygous mutants.

(A) A plant heterozygous for *Tu1-md* having *Tu1-A* allele showed half-tunicate phenotype in mature ear. **(B)** Similar ear phenotype was detected from a plant heterozygous for *Tu1-rec* that has also only one copy of *Tu1-A* allele.

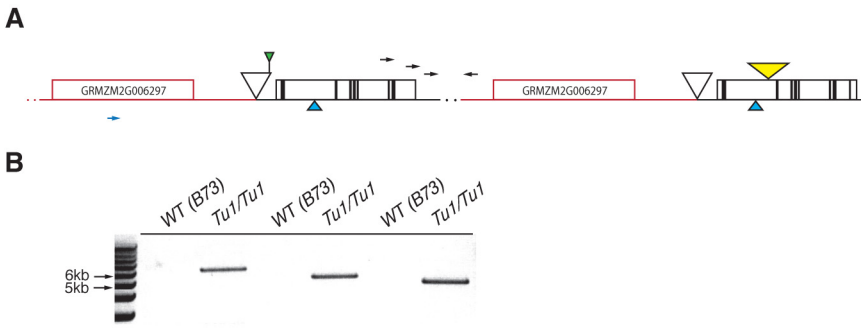


Figure 4.4. A 30 kb tandem duplication at *Tu1*.

(A) Schematic representation of duplicated *Tu1-A* and *Tu1-B* loci spanning GRMZM2G006297. Black arrows indicate the position of primers used for (**B**). **(B)** PCR showed that 3' end of *Tu1-B* is about 5 kb apart from 5' upstream of GRMZM2G006297. Sequencing and assembly suggests the presence of *gypsy-like* and *copia-like* retrotransposons and non-autonomous *dSpm-like* and *hAT-type* DNA transposons in this 5kb region. The gel image was inverted for better contrast and 1 Kb Plus DNA ladder (Invitrogen) was used to determine the size**.**

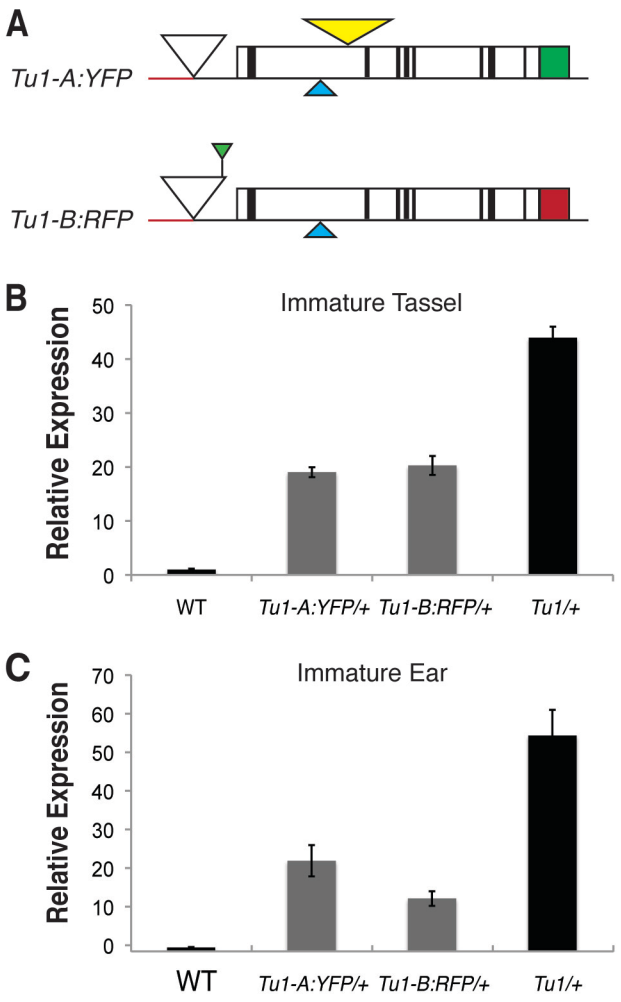


Figure 4.5. *Zmm19* expression level of *Tu1-A:YFP* and *Tu1-B:RFP* transgenic lines is correlated with copy number.

(A) Structure of *Tu1-A:YFP* and *Tu1-B:RFP* constructs. Both the 700 bp of the 3' downstream region of GRMZM2G006297 (red line) and the 2 kb novel *Mu-like* element (white triangle) were cloned with the reporter gene, *YFP* (green box) and *RFP* (red box), into the *Tu1-A* and *Tu1-B* allele and the endogenous promoter region. **(B)** The immature tassels of *Tu1-A:YFP* and *Tu1-B:RFP* transgenic lines showed half the expression level of *Zmm19* compared to heterozygous *Tu1* immature tassel (*Tu1/+*, black bar), which was used as a reference. Quantitative RT-PCR of two biological replicates. **(C)** The immature ears of *Tu1-A:YFP* and *Tu1-B:RFP* transgenic lines also showed half the expression level of heterozygous *Tu1* immature ears (*Tu1/+*, black bar). Quantitative RT-PCR of three biological replicates. Results are plotted as the ratio to the wild-type level of *Zmm19* and are represented as mean +/- SEM. *Zmm19 is* normalized to *Ubiquitin* levels. *Tu1-A:YFP/+* and *Tu1-B:RFP/+* indicate heterozygous *Tu1-A:YFP* and heterozygous *Tu1-B:RFP* plants, respectively.

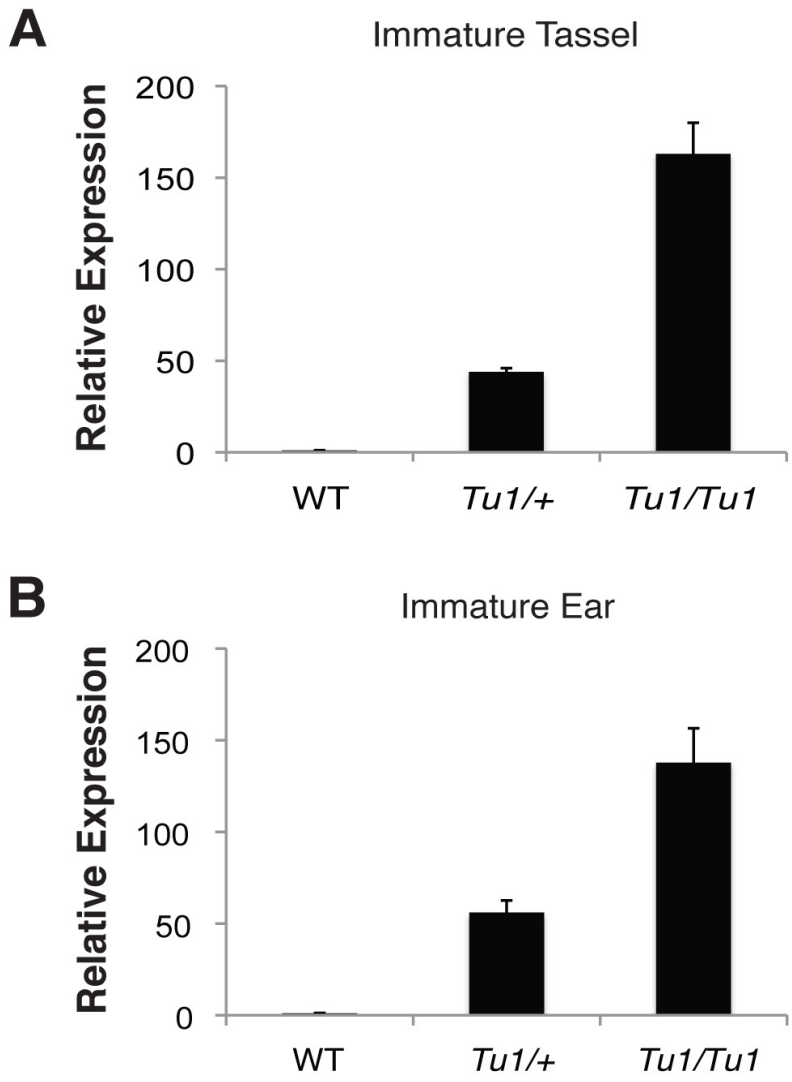


Figure 4.6. Upregulation of *Zmm19* expression in *Tu1* reproductive tissues.

(A) and **(B)** *Zmm19* expression in immature tassels (1-2 cm long) **(A)** and immature ears (0.5-0.9 cm long) **(B)** was determined by qRT-PCR. Quantitative RT-PCR of three biological replicates for each tissue. Results are plotted as the ratio to the wild-type level of *Zmm19* and are represented as mean +/- SEM. *Zmm19 is* normalized to *Ubiquitin* levels. *Tu1/+* indicates heterozygous *Tu1* plants and *Tu1/Tu1* homozygous *Tu1*.

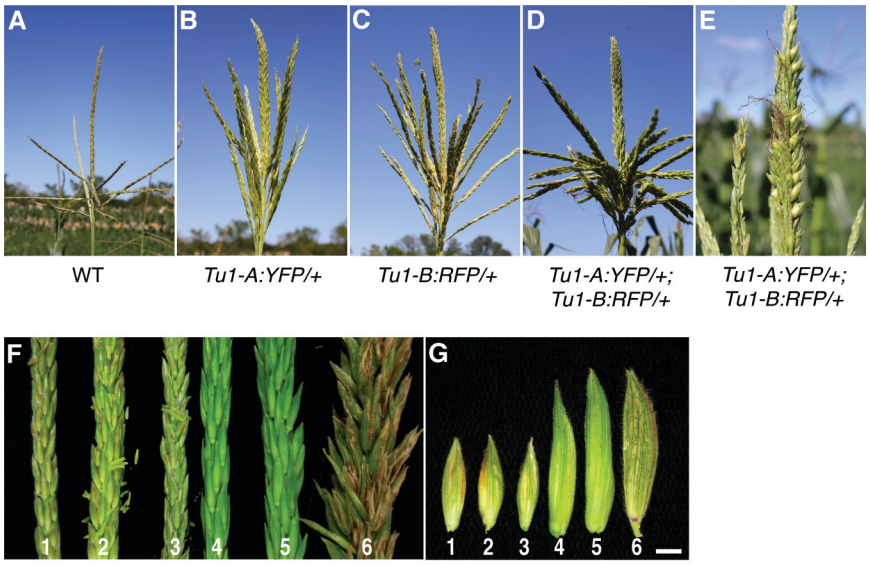


Figure 4.7. The phenotype of *Tu1-A:YFP* and *Tu1-B:RFP* transgenic plant tassels resembles that of *Tu1*.

(A) Wild-type tassel. **(B)** and **(C)** Single transgenic tassels heterozygous for *Tu1-A:YFP* **(B)** and *Tu1-B:RFP* **(C)** had a rachis and lateral branches that were relatively thicker than those of the wild-type tassel **(A)** due to elongated glumes, respectively. **(D)** and **(E)** Transgenic plants harboring two transgenes (*Tu1-A:YFP* and *Tu1-B:RFP*) showed dense spikelets with elongated glumes **(D)** and successful kernel-bearing fertilization **(E)** in the tassel. **(F)** and **(G)** *Tu1-A:YFP* and *Tu1-B:RFP* transgenic plants showed elongated glume phenotypes that are distinct from those of wild-type plants. Morphological change was detected in a single glume phenotype of the transgenic plants **(G)**, which is comparable to heterozygous *Tu1* plants. 1, B73; 2, Mo17; 3, transgenic control (WT), 4, heterozygous *Tu1-A:YFP* (*Tu1-A:YFP /+*); 5, heterozygous *Tu1-B:RFP* (*Tu1-B:RFP/+*); and 6, heterozygous *Tu1* (*Tu1/+*). Scale bar, 2.5 mm **(G)**.

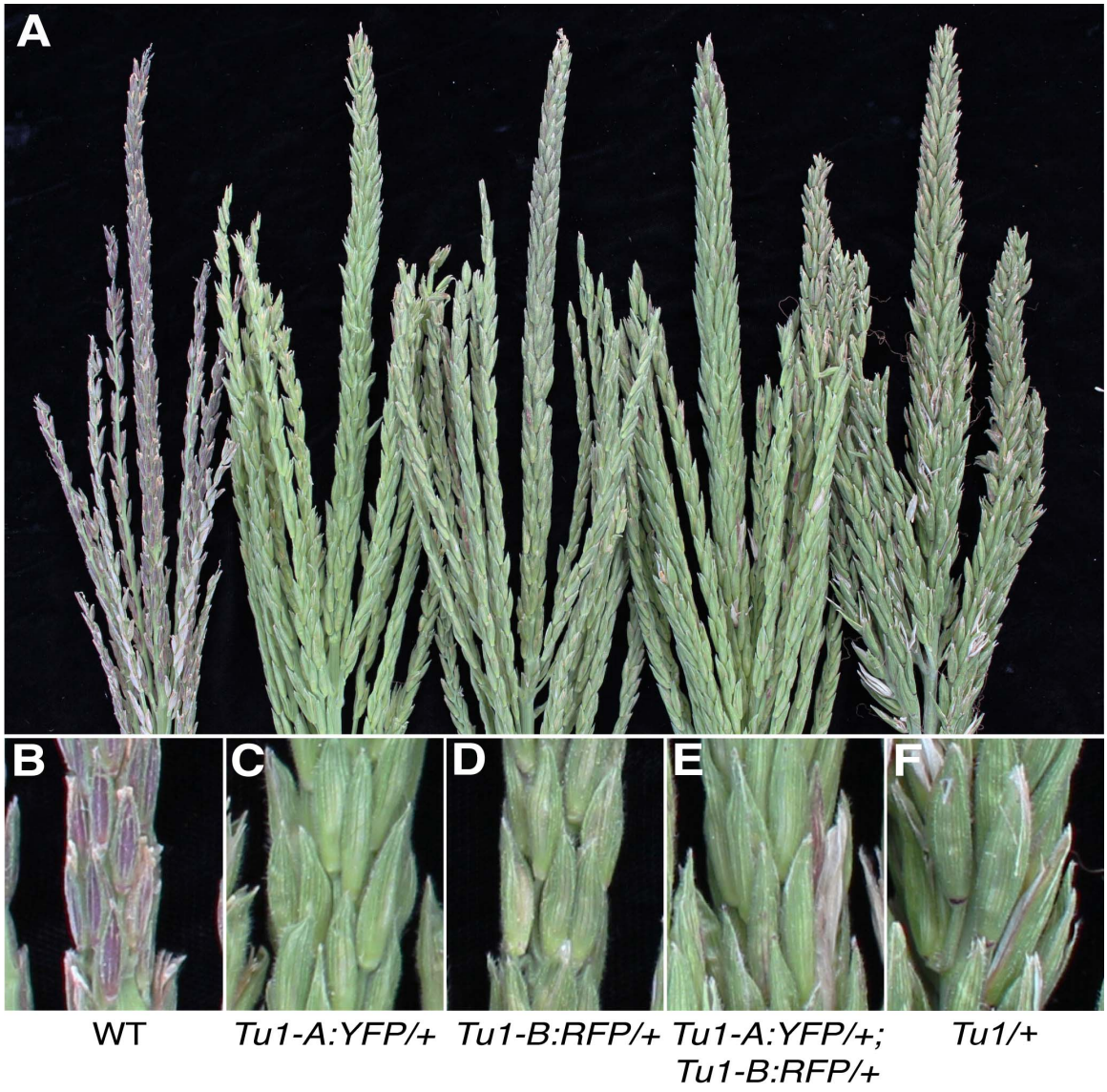


Figure 4.8. *Tu1-A:YFP* and *Tu1-B:RFP* transgenic tassels phenocopy *Tu1* in a dosage-dependent manner.

(A) T2 transgenic tassels harboring one *Tu1* transgene appear to have relatively thick rachises and lateral branches with elongated glumes. The main rachis of the double transgenic plant (*Tu1-A:YFP/+*; *Tu1-B:RFP/+*) was even thicker than that of the single transgenic plants. The order of each tassel in **(A)** corresponds to that in **(B)** to **(F)**. **(B)** to **(F)** A close-up view of **(A)**, showing the additive effect of the glume phenotype.

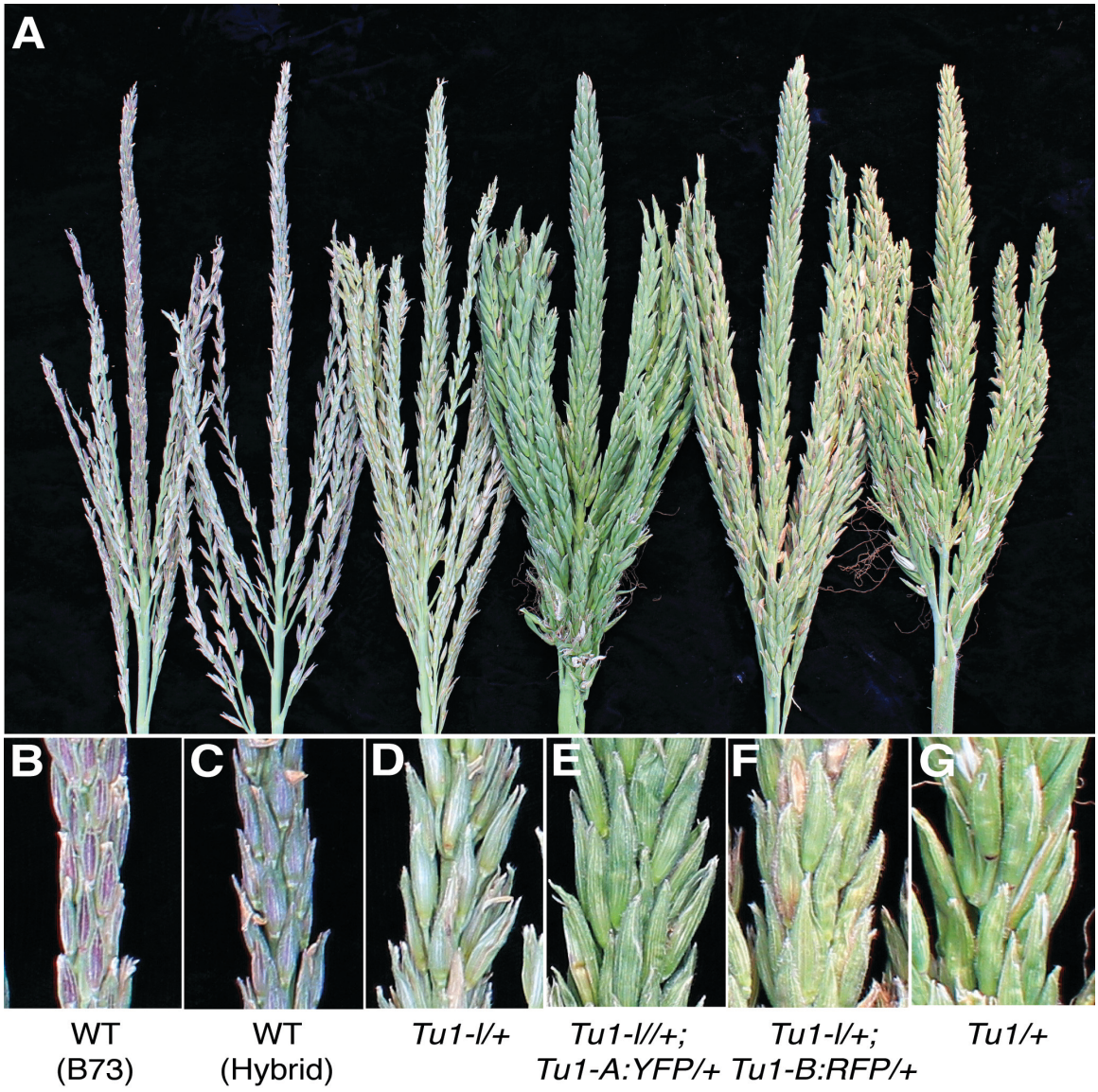


Figure 4.9. *Tu1-A:YFP* and *Tu1-B:RFP* transgenic lines interact with *Tu1-l.*

(A) Each single transgenic line produced an additive tassel phenotype with the presence of the *Tu1-l* allele compared to a single *Tu1-l*. The order of each tassel in **(A)** corresponds to those in **(B)** to **(G)**. **(B)** to **(G)** A close-up view of **(A)**, showing that glumes were elongated in a dose-dependent manner. Rachis phenotype of B73 wild-type inbred **(B)**, hybrid between B73 and the wild-type transformant **(C)**, a single-copy *Tu1-l* heterozygote **(D)**, double heterozygote for *Tu1-l* and *Tu1-A:YFP* **(E)**, double heterozygote for *Tu1-l* and *Tu1-B:RFP* **(F)**, and *Tu1* heterozygote **(G)**.

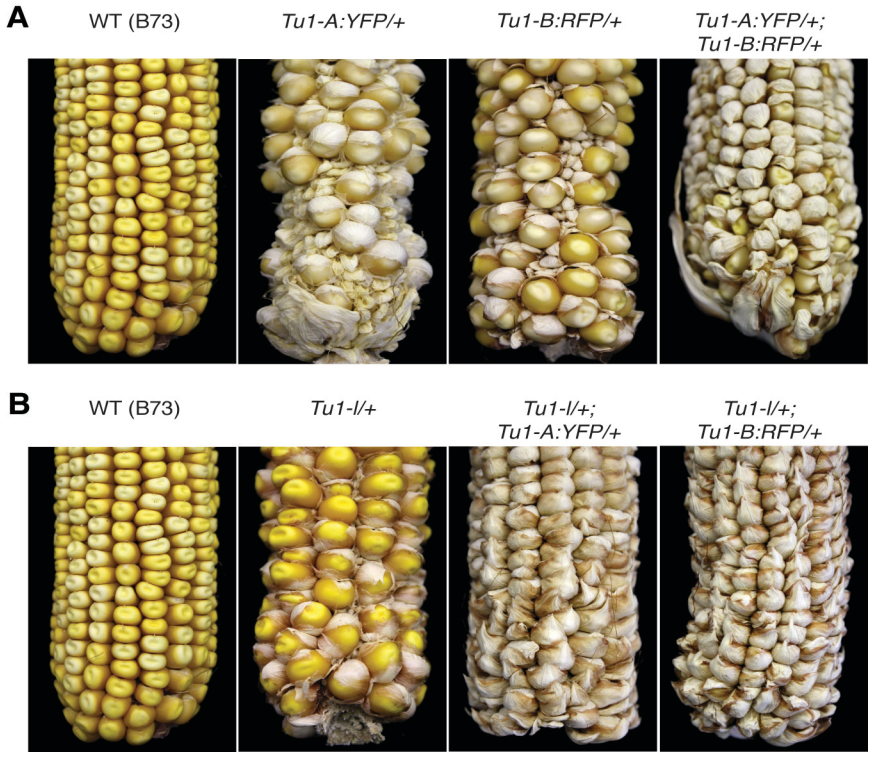


Figure 4.10. *Tu1-A:YFP* and *Tu1-B:RFP* transgenic ears phenocopy *Tu1*.

(A) The half tunicate phenotype was observed in transgenic ears carrying each single transgene (*Tu1-A:YFP/+* or *Tu1-B:RFP/+*), while the wild-type kernels were naked. Two transgenes present in one transgenic line (*Tu1-A:YFP/+*; *Tu1-B:RFP/+*) caused glumes to be further elongated to fully enclose kernels, suggesting the additive genetic effect of the two transgenes. **(B)** A single transgenic line (*Tu1-A:YFP/+* or *Tu1-B:RFP/+*) produced fully-elongated glumes with the presence of *Tu1-l,* while plants heterozygous for a single copy of *Tu1-l* (*Tu1-l/+*) represented the half tunicate phenotype that was comparable to either the *Tu1-A:YFP/+* or *Tu1-B:RFP/+* phenotype **(A)**.

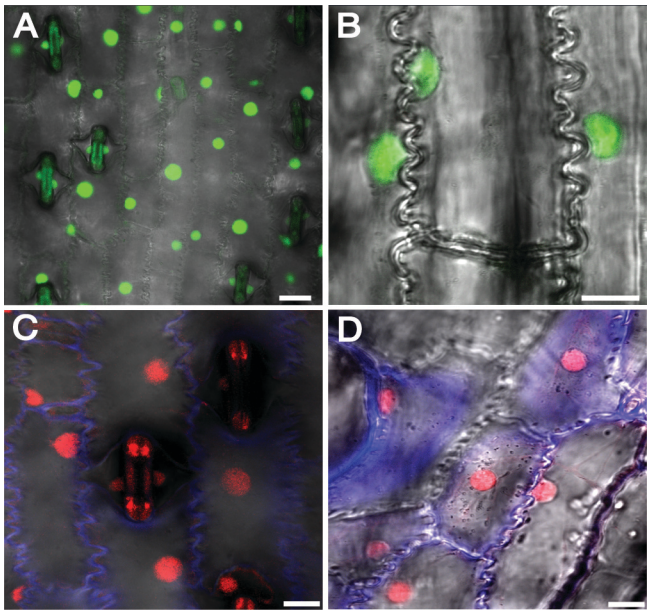


Figure 4.11. TU1-A:YFP and TU1-B:RFP are nuclear-localized.

(A) and **(B)** TU1-A:YFP fusion proteins are nuclear-localized in leaf **(A)** and glume **(B)** epidermis. **(C)** and **(D)** TU1-B:RFP fusion proteins are nuclear-localized in leaf **(C)** and glume **(D)** epidermis. Auto-fluorescence of YFP and RFP was detected from guard cells in the leaf epidermis **(A)** and **(C)**. Scale bar, 20 μm **(A)** and 10 μm **(B)** to **(D)**.

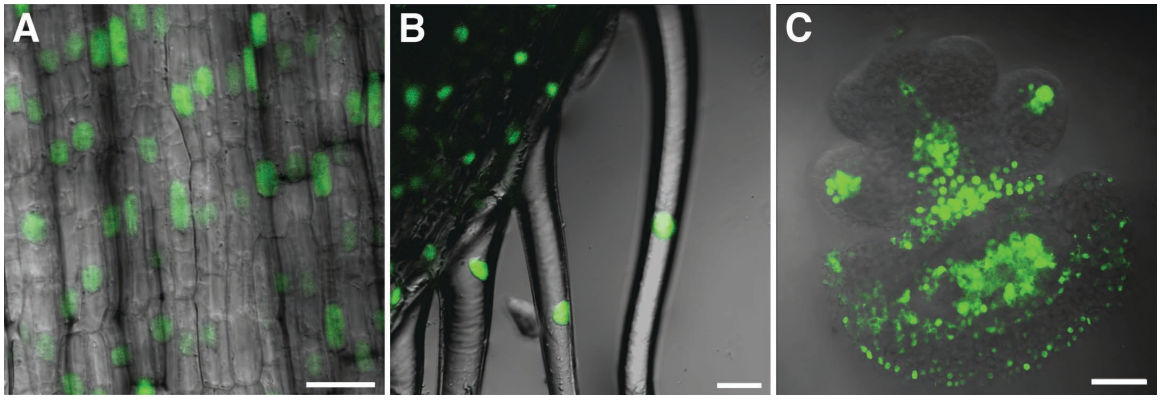


Figure 4.12. TU1-A:YFP is expressed in vegetative and reproductive tissues.

(A) to **(C)** YFP expression was clearly detected in nuclei from not only vegetative tissue, i.e., the husk **(A)**, but also reproductive tissues, i.e., trichomes **(B)**, and the floral meristem (FM, **[C]**) in 1-cm-long immature ears. Scale bar, 10 μm **(A)**, 20 μm **(B)**, and 50 μm **(C)**.

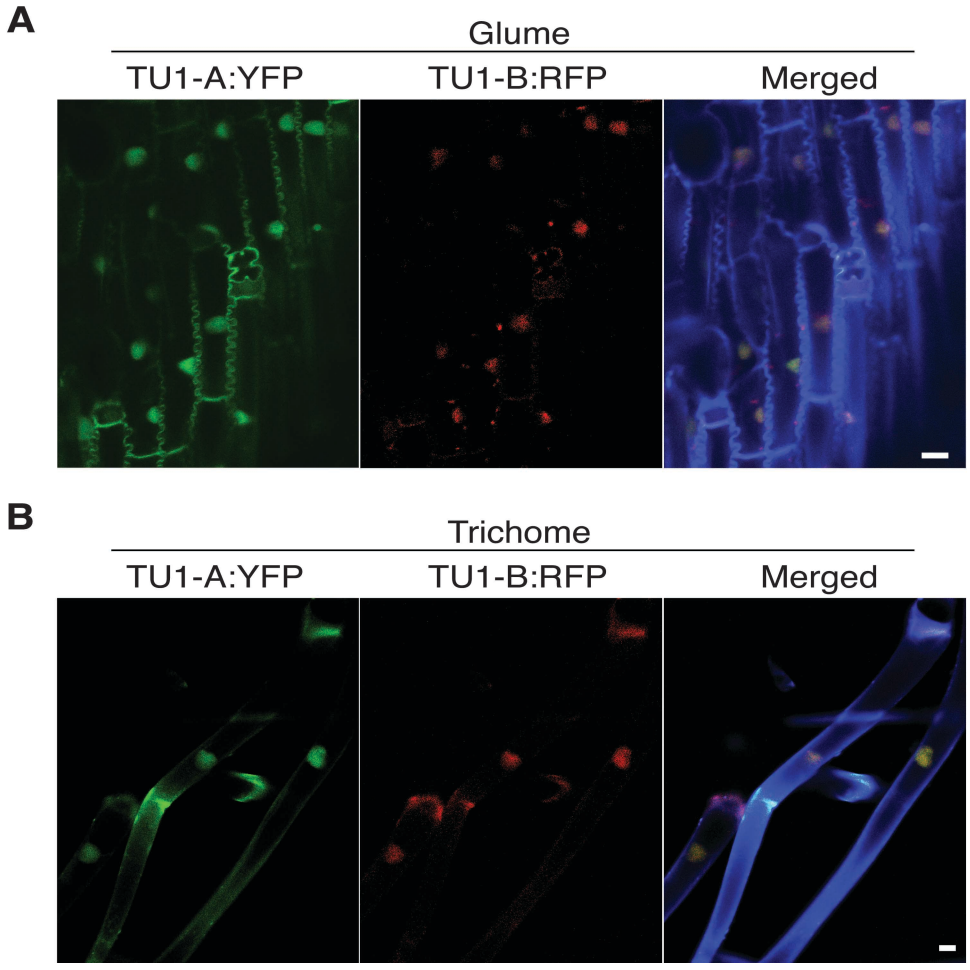


Figure 4.13. TU1-A:YFP and TU1-B:RFP proteins are co-localized in nuclei.

(A) and **(B)** Transgenic lines harboring both *Tu1-A:YFP* and *Tu1-B:RFP* displayed co-localization of TU1-A:YFP and TU1-B:RFP in the nuclei of glume **(A)** and trichome **(B)** epidermis. Merged images show that TU1-A is co-expressed with TU1-B in the same nuclei. Scale bar, 10 μm **(A)** and **(B)**.

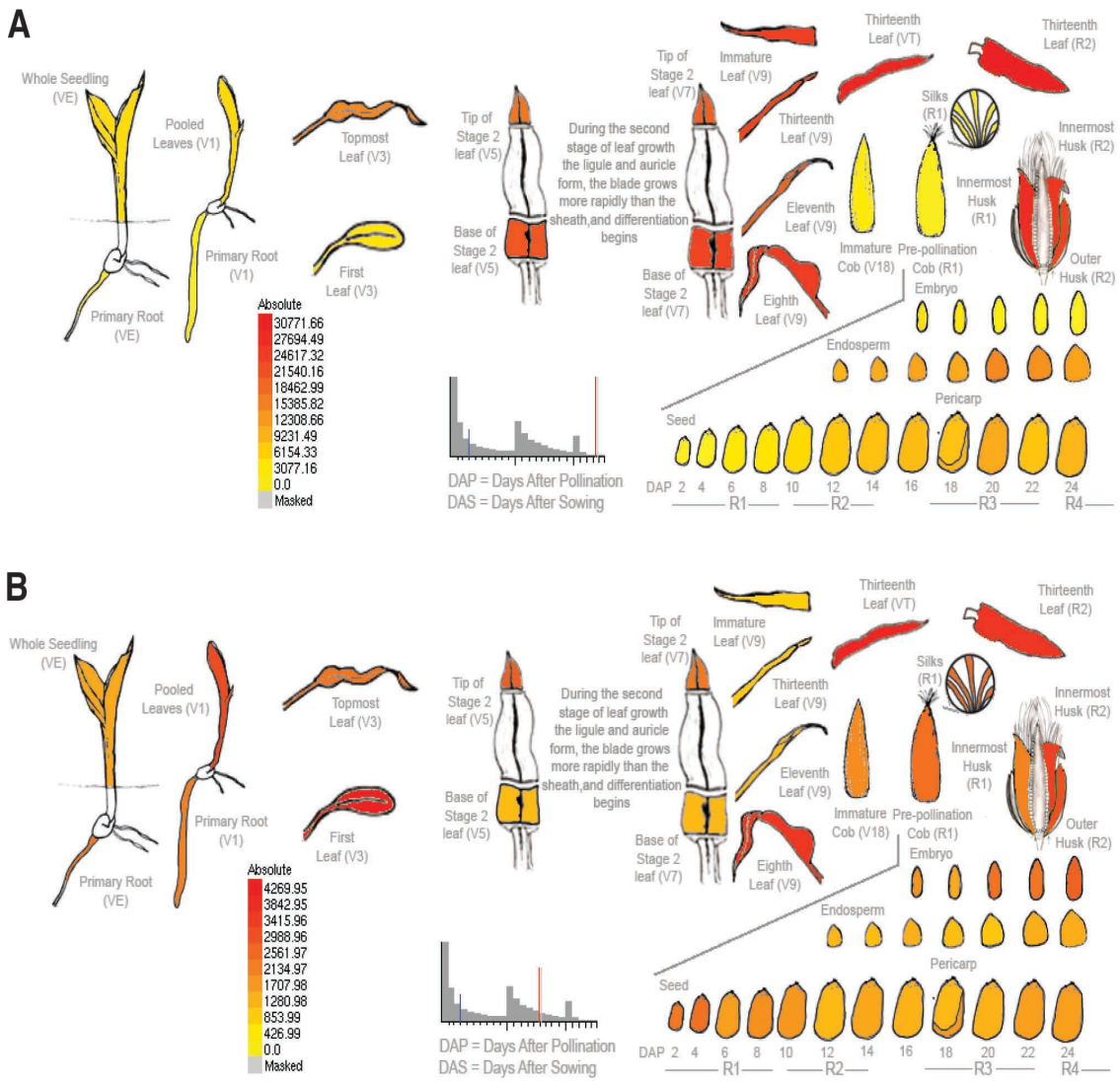


Figure 4.14. Gene expression data of GRMZM2G370777 and GRMZM2G006297.

(A) and **(B)** The maize electronic Fluorescent Pictograph Browser (eFP browser; *bar.utoronto.ca/maizeefp*) displays the expression level of target genes, GRMZM2G370777 **(A)** and GRMZM2G006297 **(B),** during maize development. The expression data are based on a report by(Sekhon et al., 2011).

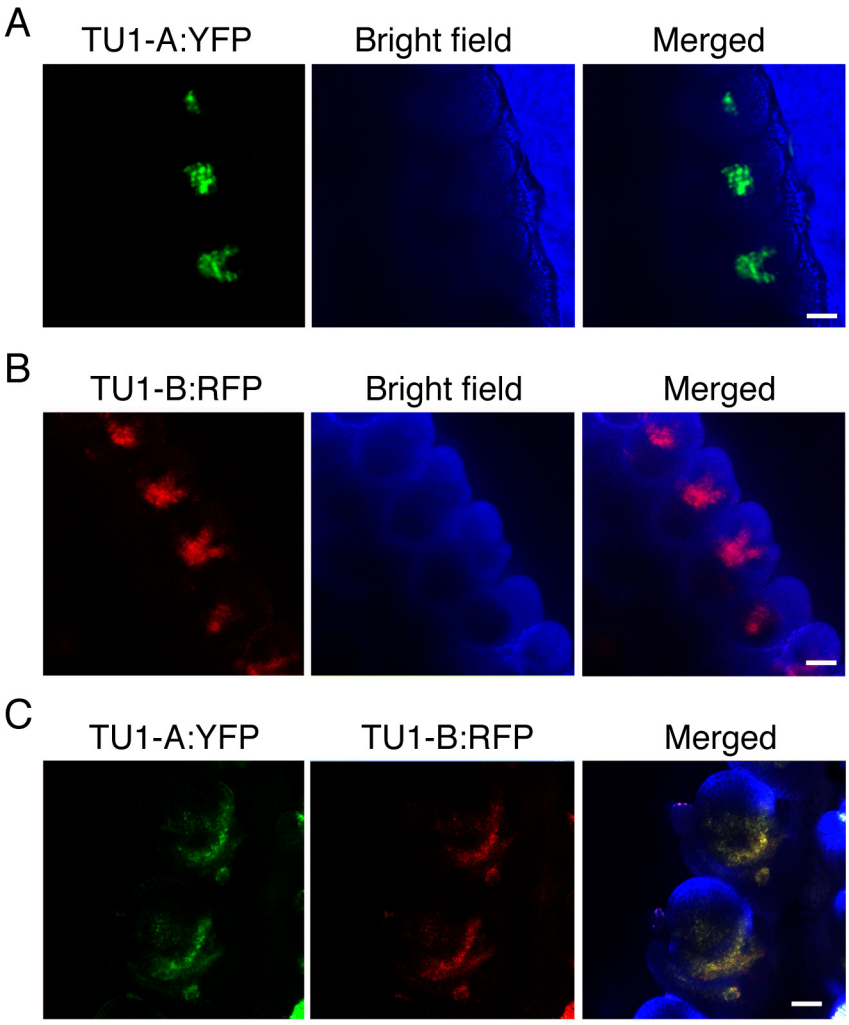
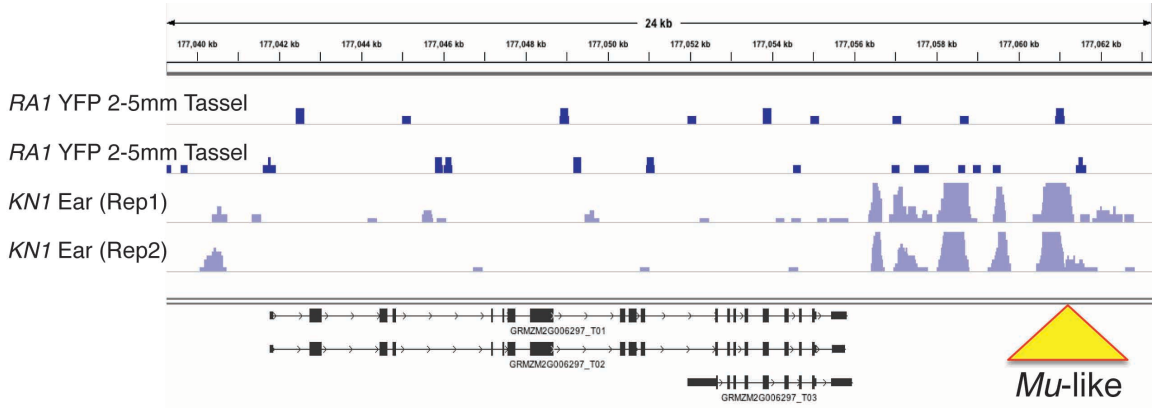


Figure 4.15. Developmental expression of TU1-A:YFP and TU1-B:RFP in immature ears.

(A) and **(B)** In 2 mm ears, TU1-A:YFP **(A)** and TU1-B:RFP **(B)** are expressed at specific cup-shaped domains subtending SPM, respectively. **(C)** Transgenic plants harboring both *Tu1-A:YFP* and *Tu1-B:RFP* showed co-localization of TU1-A:YFP and TU1-B:RFP at the base of the spikelet meristem in ~ 4-mm long ears. Bright field images were converted into blue channel for better contrast **(A)** and **(B)**. Scale bar, 20 μm **(A)** to **(C)**.

A



B

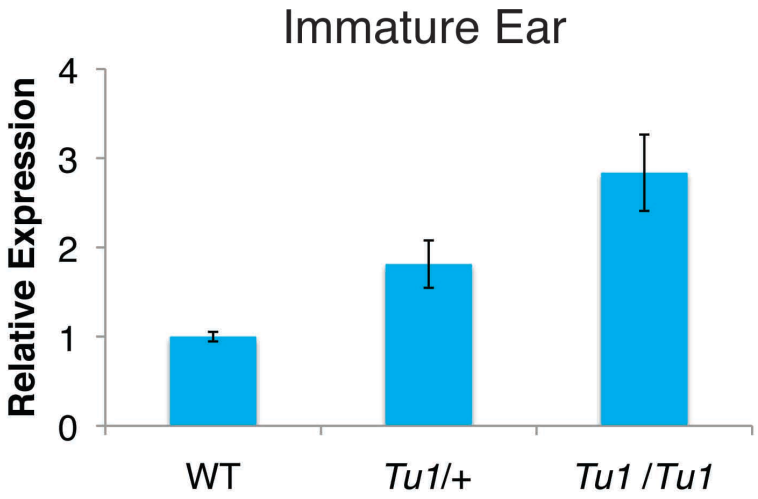


Figure 4.16. *Mu-like* insertion site in GRMZM2G006297 and upregulation of GRMZM2G006297 gene expression in *Tu1* immature ears.

(A) A screen shot of GRMZM2G006297 gene in RA1- and KN1- ChIP-sequencing data **(B)** *Zmm19* expression immature ears (0.5-0.9 cm long), which was determined by qRT-PCR. Quantitative RT-PCR of three biological replicates for each tissue. Results are plotted as the ratio to the wild-type level of *Zmm19* and are represented as mean +/- SEM. *Zmm19 is* normalized to *Ubiquitin* levels. *Tu1/+* indicates heterozygous *Tu1* plants and *Tu1/Tu1* homozygous *Tu1*.

Table 4.1. Primers used in fine-mapping and qRT-PCR.

	Forward Primer	Reverse Primer
Fine Mapping		
umc1847	GCCCAAGGTAGATTTTTACTCTCCA	AAGTCGTAGAGCTCGTCGTGATG
GRMZM2G386088	ATATTTAAGGGCACCGTGGAGAAT	CCAGACTCCCCATTGTAAAACTTG
GRMZM2G006297	GCTGCTACAGTTCGCGTTC	CGTGCTGCTTGTCAAGTTGT
IDP7898	ATGATCTCCGTGGTCTTTGG	TGCAAAGCACCTATACACGC
umc1194	ACCACCAGACATGGGAAACTTCT	AAGGCGGACACTACTCTACCCTCT
IDP8018	GAGACCTTTGACGCACTTCC	ACATGAGAAGGCGATTCTGG
GRMZM2G370777	AGATGCATTGATCGGTGATG	ACATATCCCCCACGTTGTGT
GRMZM2G081318	GTGAACCATGGCCGAGAAAC	CAGCGCACAGTGATCTAGGA
IDP4760	GCATACCCAGATTTGTGTCC	TTGCGATCACTCAGAAGACG
qRT-PCR		
Zmm19	GCTCATCGTCTTCTCCTCCA	TTGCTTCCACAAGTTGCTCA
Ubiquitin	TAAGCTGCCGATGTGCCTGCGTCG	TGAAAGACAGAACATAATGAGCACAG

Table 4.2. Primers used in transgene construction.

The att sites required for Gateway cloning are underlined.

Primer	Primer Sequence
*Tu1*AB_attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATCGCCAAGACTTTGTGAAGA
*Tu1*A_attB5(R)	GGGGACAACTTTTGTATACAAAGTTGTGATGGAGACACCGATGACTTTC
*Tu1*A_attB5	GGGGACAACTTTGTATACAAAAGTTGCAACTTGTTCGGACACTTGTCACA
*Tu1*AB_attB4	GGGGACAACTTTGTATAGAAAAGTTGGGTGCTTCCATGCAACACAAGGC
*Tu1*AB_attB3	GGGGACAACTTTGTATAATAAAGTTGCATGAAAAAAACAAGATGATCTGTGTTG
*Tu1*AB_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCTAAGGATCCAAACGATGG
*Tu1*B_attB5(R)	GGGGACAACTTTTGTATACAAAGTTGTGTAGGTCGTTTTCATGGGTTTC
*Tu1*B_attB5	GGGGACAACTTTGTATACAAAAGTTGCAACAACGTGGGGGATATGTTG
YFP_attB4(R)	GGGGACAACTTTTCTATACAAAGTTGCATAAGGCCGGCCTGGAGGTGGAGGTGG AGCT
YFP_attB3(R)	GGGGACAACTTTATATACAAAGTTGTGGCCCCAGCGGCCGCAGCAGCACCAGC AGGATC
RFP_attB4(R)	GGGGACAACTTTTCTATACAAAGTTGCAGGCCGGCCTGGAGGTGGAGGTGGAGC TGCCTCCTCCGAGGACGTCATC
RFP_attB3(R)	GGGGACAACTTTATATACAAAGTTGTGGCCCCAGCGGCCGCAGCAGCACCAGC AGGATCGGCGCCGGTGGAGTGGCGGCC

Table 4.3. Primers used in allele-specific and long-range PCRs.

	Primer Sequence	Figure
Allele-specific		
Zmm19 5' F14	TTGCCCAGGACTGACTGATA	4.2C
Tu1 5' F10	GATGCATGAAGAAGCCATGA	4.2D
Tu1 int2 R6	ACATATCCCCCACGTTGTGT	4.2C, 4.2D, 4.2E
Tu1 3' F35	TCTGTGTTGATCTGGGTGGA	4.4B
Tu1 3' F39	CGTCGAAGCTATCGATTATGAAG	4.4B
2G006297 R11	CACTCTGATACCCAGCAATCG	4.4B
Long-range		
2G006297 F44	GCTTCCATGGTTGACCTTCT	4.2E
Tu1 Ex6 F34	GGAGAACAGGCAACTGAGGA	4.2F, 4.4B
2G006297 R21	AAGCAGTCGATTCCTCCTCA	4.2F

Methods

Plant Material

The *Tu1* (No. 412G)*, Tu1-l* (No. 416B)*,* and *Tu1-md* (No. 416E) alleles in an unknown genetic background were obtained from the Maize Genetics Coop Stock Center. *Tu1*, *Tu1-l, and Tu1-md* plants were introgressed into B73 two times. B73 was used as the wild-type line. Plants were grown in the field or in the green house under standard conditions.

Fine Mapping

Heterozygous *Tu1* plants were crossed to the B73 inbred line and then F1 heterozygous *Tu1* plants were backcrossed to B73 to generate mapping populations segregating equally for wild type and mutant at a 1:1 ratio. Phenotypes of ~750 F2 plants were scored by visual inspection of mature tassels and ears. DNA preps were done on 738 plants for positional cloning. Simple Sequence Repeat (SSR), InDel Polymorphism (IDP) markers were used on Chromosome 4L, where the *Tu1* locus is located, as previously described (Mangelsdorf and Galinat, 1964). The F2 population was screened with markers IDP8954 and umc2009 on chromosome 4L to identify recombinants. To narrow the mapping interval, Cleaved amplified polymorphic sequence (CAPS) (Konieczny and Ausubel, 1993) and Insertion/Deletion (InDel) markers were designed for maize genes by sequencing and identifying sequence polymorphisms between B73 and the *Tu1* progenitor.

Genetic markers were amplified with 20 – 60 ng of DNA and Phusion High-Fidelity DNA polymerase (FINNZYMES) by PCR using primers that are listed in Table 4.1. SSR and IDP markers were tested according to the recommended PCR conditions (http://www.maizegdb.org/documentation/maizemap/ssr_protocols.php) and (magi.plantgenomics.iastate.edu/browseMarker.do), respectively. As a CAPS marker, GRMZM2G386088 PCR product was subsequently treated with the KpnI restriction enzyme at 37°C for 1 hour and analyzed via 1% agarose gel electrophoresis.

Scanning Electron Microscopy

As described in Whipple et al. (2010), fresh samples of maize immature ears were dissected and mounted on disks with silver adhesive (Electron Microscopy Sciences) and placed on ice before imaging. A Hitachi S-3500N scanning electron microscope was used to capture images of the live sample by the electron beam using an accelerating voltage of 5.0 kV under high vacuum mode and a distance of 15 to 30 mm within 15 minutes.

Tu1-A:YFP and *Tu1-B:RFP* transgenic maize lines

Tu1-A:YFP and *Tu1-B:RFP* transgenes were constructed by cloning *Tu1-A* and *Tu1-B* alleles using homozygous *Tu1* DNA, respectively. *Tu1-A:YFP* construct was generated by fusing YFP in-frame to the C-terminus before the stop codon of the genomic sequence of *Tu1-A,* including 700 bp of the 3' downstream region of GRMZM2G006297, 2 kb of the *Mu-like* element, 300 bp of the 5' promoter, a complete *Tu1-A* coding sequence with a 3.5 kb insertion in the first intron, and 450 bp of the 3'UTR. The MultiSite Gateway Four Fragment System (Invitrogen) was used, by modifying the method described in Mohanty et

al. (2009). All fragments were amplified using Phusion *Taq* polymerase (FINNZYMES). The first fragment from the 5' upstream region to the 3.5-kb insertion within the first intron was amplified with primers *Tu1*AB_attB1 and *Tu1*A_attB5(R) and cloned into the pDONR221 P1-P5R vector using BP recombinase (Invitrogen). The second fragment from the remaining first intron to the coding sequence before the stop codon was amplified with primers *Tu1*A_attB5 and *Tu1*AB_attB4 and cloned into the pDONR221 P5-P4 vector. The citrine YFP fragment was amplified with primers YFP_attB4(R) and YFP_attB3(R) and cloned into the pDONR221 P4r-P3r vector using BP recombinase (Invitrogen). The third fragment from the stop codon to the 3'UTR was amplified with primers *Tu1*AB_attB3 and *Tu1*AB_attB2 and cloned into the pDONR221 P3-P2 vector. The pDONR221 P1-P5R, pDONR221 P5-P4, pDONR221 P4r-P3r, and pDONR221 P3-P2 vector fragments were combined and transferred to the pTF101 Gateway compatible maize transformation vector by a multisite LR recombination reaction (Invitrogen). Confirmed clones were transferred to *Agrobacterium* and transformed into maize (Mohanty et al., 2009). The same protocol was applied for the *Tu1-B:RFP* construction. For the *Tu1-B:RFP* construct, *Tu1*B_attB5(R) and *Tu1*B_attB5 primers were used to amplify the first and second fragment specific to *Tu1*-B. The mRFP1 fragment was amplified with primer RFP_attB4(R) and RFP_attB3(R) and cloned into the pDONR221 P4r-P3r vector. The same third fragment was used because the 3'UTR is identical between *Tu1-A* and *Tu1-B.* Likewise, four vector fragments were combined and transferred to the pTF101 vector. Confirmed clones were transferred to *Agrobacterium* and transformed into maize. The sequences of the primers are shown in Table 4.2.

Allele-specific, Long-range PCRs and Sequencing

For all PCR reactions, genomic DNA was extracted from wild-type inbred lines and plants homozygous for *Tu1* mutations. PCR amplification of *Tu1* alleles was performed with Phusion High-Fidelity DNA polymerase (FINNZYMES) and allele-specific primers (Table 4.3) under optimal PCR conditions, according to Finnzymes' recommendations. For the long-range PCRs, TaKaRa LA *Taq* polymerase was used with primers (Table 4.3) under two-step PCR conditions following the provided protocol. The PCR product was subsequently prepared for Illumina DNA sequencing according to the manufacturer's recommendations. Sequencing was performed on an Illumina Genome Analyzer "GAII" for PE50 cycles.

qRT-PCR

Total RNA was extracted from immature tassels (1-2 cm long) and immature ears (0.5-0.9 cm long) using TRIzol reagent (Invitrogen). Total RNA was treated with DNaseI and reverse transcribed using an oligo-dT primer and SuperScript III Reverse Transcriptase (Invitrogen). qRT-PCR was performed with iQ SYBR Green Supermix (BioRad Laboratories) using two technical replicates each of two or three biological replicates. qRT-PCR primers are listed in Table 4.1. qPCR reactions were annealed at 57°C. The relative expression values for all experiments were calculated based on the expression of the experimentally validated control gene *Ubiquitin* as previously described (Satoh-Nagasawa et al., 2006). qPCR was performed on a CFX96 thermocycler and the results analyzed on the CFX Manager Software package (BioRad Laboratories). Relative expression was calculated using the 'delta-delta method' formula $2^{-[\Delta CP\ sample - \Delta CP\ control]}$, where 2 represents perfect

PCR efficiency.

Confocal Microscopy

Inflorescences in different developmental stages were hand-sectioned and visualized in water. Immature ears were counter-stained with calcoflour for 2 minutes and washed with a buffer (50% glycerol, 1X PBS, and 0.1% NaAzide). Fluorescent proteins were detected on an LSM 710 confocal microscope (Carl Zeiss). Bright field images were collected together with fluorescence images using the transmitted light detector and were processed into a blue background using Zeiss ZEN software and Adobe Photoshop CS4 (Adobe Systems).

Accession numbers

Sequence data from this article can be found in the GenBank and MaizeSequence (MaizeSequence.org) databases under the following accession numbers: *Tu1-A* (AJ850302), *Tu1-B* (AJ850303), *Tu1-d* (AJ850299), *Tu1-l* (AJ850300), and *Tu1-md* (AJ850301), GRMZM2G370777, and GRMZM2G006297.

Chapter V: Concluding Remarks

Conclusions

Transposable elements have been viewed as genomic parasites that survive by their capability to amplify faster than they are deleted by their hosts. Diversification of TEs has been adopted as one of their defense mechanisms to prosper over a long evolutionary time. Ironically, the mobile behavior of TEs has imposed on the hosts different levels of pressure to evolve gene regulation systems and to reconstruct chromosomal composition. Consequently, transposons have become a source of material for coordination of eukaryotic gene regulatory systems transcriptionally and post-transcriptionally.

One of DNA transposons, *Mutator,* in maize has been utilized for insertional mutagenesis for the purpose of saturating the maize genome with new alleles in each gene. The progress on identifying flanking sequence tags (FSTs) from two *Mu* insertional collections (*RescuMu* and *UniformMu*) has been slow, as there were no reliable methods to isolate flanking sequences in a large-scale. We have sought to develop a method for efficient and large-scale FST identification in MTM lines, using a modified ligation-mediated GenomeWalker protocol coupled with high-throughput DNA sequencing technology (Chapter 2). The advantage of this system is that we could maximize the recovery of FSTs by the GenomeWalker method to enrichment flanking sequences and the illumina genomic sequencing to provide enough read lengths and depth for capturing less frequent germinal insertions. In terms of an inter-chromosomal view, MTM *Mu* insertion

sites are largely distributed all over ten chromosomes and the number is relatively proportional to each chromosome size. However, in a view of an intra-chromosomal distribution, unique germinal insertions (UGIs) tend to target in a non-uniform manner, favoring the arms of each chromosome and avoiding the centromere(s). Hypomehtylation of *Mu* flanking regions is also supportive to explain 'U-shaped pattern' of the non-uniform Mu distribution because genome-wide DNA methylation levels in CG and CHG contexts are decreased toward the arms of chromosomes (Liu et al., 2009; Gent et al., 2012). By non-overlapping parallel mapping strategy, we have characterized 572 parental insertion sites. On the other hand, we have found in average 162 unique germinal insertions in a single plant by cross-reference mapping strategy. As we have sequenced 96 libraries from 48 column and row pools, the completion of *Mu* mapping will allow us to create sequence-indexed *Mu* FST database and we expect to find about 370,000 germinal insertion candidates from one grid if each plant bears 162 new germinal insertions, and we successfully validated several parental and germinal *Mu* insertions via PCR validation assays. If we assume that 162 *Mu* copies are transmitted to F1 generation, 410 parental insertions must have been immobilized. It is possible that we obtained insertion sites of immovable Pack-MULEs that share TIRs with high sequence homology.

Another DNA transposons, *Ac/Ds,* in maize have been utilized for regional and reconstitutional mutagenesis for saturating linked sites with multiple new alleles. Beside the utility of Ac/Ds system, a striking phenomenon, which is negative dosage effect, has been observed for decades. To address whether or not small RNAs play a role in negative dosage effect of *Ac*, we have investigated transcriptional abundance of *Ac* and *Ds* elements and the presence/absence of small RNA mapped to *Ac* and *Ds* elements (Chapter 3). What we have

found is that both sense *Ac* transcript and antisense *Ac* fusion transcript are more abundant in endosperms carrying three copies of *Ac-im* than in those carrying a single copy of *Ac-im.* This result suggests the possibility of double-stranded RNA (dsRNA) but we did not collect any experimental evidence of dsRNA formation between sense *Ac* and antisense *Ac* transcripts by far. Instead of concluding no dsRNA formation, we open up a possibility of insufficient conduct on experimental procedure for dsRNA detection assay. Interestingly, small RNA sequencing data presented that 21 nt siRNA derived from *Ds* element not *Ac-im* was detected at the border of 5' and 3' end of *Ds*6-like element, which relies on *Ac-im* copy number. Presumably, *Ac-Ds* dsRNA formation would take place in endosperms carrying *Ac-im* copi(es) to produce *Ac-im* dependent *Ds*-derived 21 nt siRNA. From 22 nt siRNA profile, we have detected a peak within 5' subterminal region mainly in endosperms carrying a single copy of *Ac-im*, yet more detailed investigation will be necessary to understand the role of 22 nt siRNA. Finally, we have found that 24 nt siRNA are detected at the 5' and 3' subterminal regions of *Ac-im* and *Ds* in all series of *Ac-im* dosage, suggesting that 24 nt sRNA are not associated with *Ac* dosage effect. Base on our mRNA and siRNA expression data, our results do not support the idea of association of small RNA with negative dosage effect.

Lastly, DNA transposons can cause chromosomal rearrangement that reconstruct genome structure (Bennetzen, 2005; Bourque, 2009; Parisod et al., 2010). Sometimes, alternative transposition of DNA transposons induces a partial cut-and-paste activity, resulting in chromosomal deletions, inversions, duplications, and translocations (Gray, 2000; Lim and Simmons, 1994; Zhang and Peterson, 2004). From our study on a maize classical *Tunicate1* mutant, exhibiting pod corn (Zea mays var tunicata) phenotype, we have

characterized that a novel *Mu-like* element gave rise to the rearrangement of *Tunicate1* (*Tu1*) locus (Chapter 4). Our study revealed that *Tu1* is a dominant pod corn mutation in which kernels are completely enclosed in leaf-like glumes. We have shown that *Tu1* encodes a MADS box transcription factor expressed in leaves whose 5' regulatory region is fused by a 1.8-Mb chromosomal inversion to the 3' region of a gene expressed in the inflorescence. Interestingly, both genes, *Zmm19* and the upstream gene (GRMZM2G006297), are further duplicated, resulting in dosage-dependent upregulation of both genes. Furthermore, *Tu1* transgenes phenocopied the pod corn and interacted in a dose-dependent manner with the classical derivative alleles resulting from chromosomal rearrangement. We have also found that TU1 proteins are nuclearly localized in specific cells at the base of spikelet pair meristems in young ear primordia. *Tu1* branch determination defects resemble those in ramosa mutants, which encode regulatory proteins expressed in these same cells, suggesting that *Tu1* is involved in inflorescence branching. Taken together, transposon-mediated chromosomal rearrangement and subsequent regional duplication at *Tu1* locus influence not only structural change of the genome but also transcriptional regulatory networks of adjacent genes.

Thus, our work supports that the movement and accumulation of DNA transposons have been a significant feature for insertional mutagenesis to harvest new alleles, chromosomal reconstruction to revise a genetic/genomic map, and regulatory control to coordinate gene networks.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., *et al.* (2000). The genome sequence of Drosophila melanogaster. Science **287**, 2185-2195.

Alleman, M., and Freeling, M. (1986). The Mu transposable elements of maize: evidence for transposition and copy number regulation during development. Genetics **112**, 107-119.

Alleman, M., and Kermicle, J.L. (1993). Somatic Variegation and Germinal Mutability Reflect the Position of Transposable Element Dissociation within the Maize R-Gene. Genetics **135**, 189-203.

Alonso, J.M. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana (vol 301, pg 653, 2003). Science **301**, 1849-1849.

Aufsatz, W., Mette, M.F., Matzke, A.J.M., and Matzke, M. (2004). The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. Plant Molecular Biology **54**, 793-804.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell **15**, 2730-2741.

Bai, L., Singh, M., Pitt, L., Sweeney, M., and Brutnell, T.P. (2007). Generating novel allelic variation through Activator insertional mutagenesis in maize. Genetics **175**, 981-992.

Banks, J.A., Masson, P., and Fedoroff, N. (1988). Molecular Mechanisms in the Developmental Regulation of the Maize Suppressor-Mutator Transposable Element. Gene Dev **2**, 1364-1380.

Barkan, A., and Martienssen, R.A. (1991). Inactivation of maize transposon Mu suppresses a mutant phenotype by activating an outward-reading promoter near the end of Mu1. Proc Natl Acad Sci U S A **88**, 3502-3506.

Becker, H.A., and Kunze, R. (1997). Maize Activator transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. Molecular & General Genetics **254**, 219-230.

Bennetzen, J.L. (1984). Transposable element Mu1 is found in multiple copies only in Robertson's Mutator maize lines. J. Mol. Appl. Genet. **2**, 519-524.

Bennetzen, J.L. (1996). The Mutator transposable element system of maize. Curr Top Microbiol Immunol **204**, 195-229.

Bennetzen, J.L. (2005). Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev **15**, 621-627.

Bennetzen, J.L., and Springer, P.S. (1994). The Generation of Mutator Transposable Element Subfamilies in Maize. Theoretical and Applied Genetics **87**, 657-667.

Bennetzen, J.L., Springer, P.S., Cresse, A.D., and Hendrickx, M. (1993). Specificity and Regulation of the Mutator Transposable Element System in Maize. Crit Rev Plant Sci **12**, 57-95.

Bensen, R.J., Johal, G.S., Crane, V.C., Tossberg, J.T., Schnable, P.S., Meeley, R.B., and Briggs, S.P. (1995). Cloning and characterization of the maize An1 gene. Plant Cell **7**, 75-84.

Bolduc, N., and Hake, S. (2009). The Maize Transcription Factor KNOTTED1 Directly Regulates the Gibberellin Catabolism Gene ga2ox1. Plant Cell **21**, 1647-1658.

Bolduc, N., Yilmaz, A., Mejia-Guerra, M.K., Morohashi, K., O'Connor, D., Grotewold, E., and Hake, S. (2012). Unraveling the KNOTTED1 regulatory network in maize meristems. *Submitted*.

Bortiri, E., Chuck, G., Vollbrecht, E., Rocheford, T., Martienssen, R., and Hake, S. (2006). ramosa2 encodes a LATERAL ORGAN BOUNDARY domain protein that determines the fate of stem cells in branch meristems of maize. Plant Cell **18**, 574-585.

Bortiri, E., and Hake, S. (2007). Flowering and determinacy in maize. J. Exp. Bot. **58**, 909-916.

Bourque, G. (2009). Transposable elements in gene regulation and in the evolution of vertebrate genomes. Current Opinion in Genetics & Development **19**, 607-612.

Brown, J., and Sundaresan, V. (1992). Genetic-Study of the Loss and Restoration of Mutator Transposon Activity in Maize - Evidence against Dominant-Negative Regulator Associated with Loss of Activity. Genetics **130**, 889-898.

Brutnell, T.P., May, B.P., and Dellaporta, S.L. (1997). The Ac-st2 element of maize exhibits a positive dosage effect and epigenetic regulation. Genetics **147**, 823-834.

Bundock, P., and Hooykaas, P. (2005). An Arabidopsis hAT-like transposase is essential for plant development. Nature **436**, 282-284.

Caldwell, E.E.O., and Peterson, P.A. (1992). The Ac and Uq Transposable Element Systems in Maize - Interactions among Components. Genetics **131**, 723-731.

Calvi, B.R., Hong, T.J., Findley, S.D., and Gelbart, W.M. (1991). Evidence for a Common Evolutionary Origin of Inverted Repeat Transposons in Drosophila and Plants - Hobo, Activator, and Tam3. Cell **66**, 465-471.

Candela, H., and Hake, S. (2008). The art and design of genetic screens: maize. Nat Rev Genet **9**, 192-203.

Carpenter, A.E., and Sabatini, D.M. (2004). Systematic genome-wide screens of gene function. Nat Rev Genet **5**, 11-22.

Chandler, V.L., and Hardeman, K.J. (1992). The Mu Elements of Zea-Mays. Adv Genet **30**, 77-122.

Chandler, V.L., and Walbot, V. (1986). DNA modification of a maize transposable element correlates with loss of activity. Proc Natl Acad Sci U S A **83**, 1767-1771.

Chen, H.M., Chen, L.T., Patel, K., Li, Y.H., Baulcombe, D.C., and Wu, S.H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Natl Acad Sci U S A **107**, 15269-15274.

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science **303**, 2022-2025.

Chia, J.-M., Song , C., Bradbury, P.J., Costich, D., de Leon, N., Doebley, J., Elshire, R.J., Gaut, B., Geller, L., Glaubitz, J.C., *et al.* (2012). Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet., *Accepted*.

Chomet, P., Lisch, D., Hardeman, K.J., Chandler, V.L., and Freeling, M. (1991). Identification of a regulatory transposon that controls the Mutator transposable element system in maize. Genetics **129**, 261-270.

Chomet, P.S., Wessler, S., and Dellaporta, S.L. (1987). Inactivation of the maize transposable element Activator (Ac) is associated with its DNA modification. EMBO J **6**, 295-302.

Chuck, G., Cigan, A.M., Saeteurn, K., and Hake, S. (2007a). The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. Nat. Genet. **39**, 544-549.

Chuck, G., Meeley, R., Irish, E., Sakai, H., and Hake, S. (2007b). The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. Nat. Genet. **39**, 1517-1521.

Collins, G.N. (1917). Hybrids of Zea tunicata and Zea ramosa. Proc. Natl. Acad. Sci. USA **3**, 345-349.

Conrad, L.J., and Brutnell, T.P. (2005). Ac-Immobilized, a stable source of Activator transposase that mediates sporophytic and gametophytic excision of Dissociation elements in maize. Genetics **171**, 1999-2012.

Coupland, G., Baker, B., Schell, J., and Starlinger, P. (1988). Characterization of the maize transposable element Ac by internal deletions. EMBO J **7**, 3653-3659.

Cowan, R.K., Hoen, D.R., Schoen, D.J., and Bureau, T.E. (2005). MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. Molecular Biology and Evolution **22**, 2084-2089.

Cowperthwaite, M., Park, W., Xu, Z.N., Yan, X.H., 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.

Daboussi, M.J., and Capy, P. (2003). Transposable elements in filamentous fungi. Annu Rev Microbiol **57**, 275-299.

Das, L., and Martienssen, R. (1995). Site-Selected Transposon Mutagenesis at the Hcf106 Locus in Maize. Plant Cell **7**, 287-294.

de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., and Cons, I.H.G.S. (2001). Initial sequencing and analysis of the human genome (vol 409, pg 860, 2001). Nature **412**, 565-566.

Dempsey, E. (1993). The Ac2-Bz2m Mutable System of Maize. Maydica **38**, 151-161.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nature Genetics **35**, 41-48.

Dooner, H.K., and Belachew, A. (1989). Transposition Pattern of the Maize Element Ac from the Bz-M2(Ac) Allele. Genetics **122**, 447-457.

Dooner, H.K., Robbins, T.P., and Jorgensen, R.A. (1991). Genetic and developmental control of anthocyanin biogenesis. Annu. Rev. Genet. **25**, 173-199.

Dooner, H.K., and Weill, C.F. (2007). Give-and-take: interactions between DNA transposons and their host plant genomes. Current Opinion in Genetics & Development **17**, 486-492.

Evgen'ev, M.B., and Arkhipova, I.R. (2005). Penelope-like elements - a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenet Genome Res **110**, 510-521.

Eyster, W.H. (1921). The Linkage Relations between the Factors for Tunicate Ear and Starchy-Sugary Endosperm in Maize. Genetics **6**, 209-240.

Fernandes, J., Dong, Q.F., 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**.

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet **41**, 331-368.

Feschotte, C., and Wessler, S.R. (2001). Treasures in the attic: Rolling circle transposons discovered in eukaryotic genomes. P Natl Acad Sci USA **98**, 8923-8924.

Finnegan, D.J. (1989). Eukaryotic Transposable Elements and Genome Evolution. Trends in Genetics **5**, 103-107.

Fu, H.H., Park, W.K., Yan, X.H., Zheng, Z.W., Shen, B.Z., and Dooner, H.K. (2001). The highly recombinogenic bz locus lies in an unusually gene-rich region of the maize genome. P Natl Acad Sci USA **98**, 8903-8908.

Fusswinkel, H., Schein, S., Courage, U., Starlinger, P., and R., K. (1991). Detection and abundance of mRNA and protein encoded by transposable element activator (Ac) in maize. Mol. Gen. Genet. **225**, 186-192.

Gent, J.I., Dong, Y.Z., Jiang, J.M., and Dawe, R.K. (2012). Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. Nucleic Acids Res **40**, 1550-1560.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C.J., Du, T.T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z.P., *et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science **320**, 1077-1081.

Goodwin, T.J.D., and Poulter, R.T.M. (2001). The DIRS1 group of retrotransposons. Molecular Biology and Evolution **18**, 2067-2082.

Goodwin, T.J.D., and Poulter, R.T.M. (2004). A new group of tyrosine recombinase-encoding retrotransposons. Molecular Biology and Evolution **21**, 746-759.

Grant, S.R., Gierl, A., and Saedler, H. (1990). En/Spm encoded TnaA protein required a specific target sequence for suppression. EMBO J **9**, 2029-2035.

Gray, Y.H. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet **16**, 461-468.

Greenblatt, I.M., and Brink, R.A. (1962). Twin Mutations in Medium Variegated Pericarp Maize. Genetics **47**, 489-501.

Grotewold, E., Athma, P., and Peterson, T. (1991). A Possible Hot-Spot for Ac Insertion in the Maize P-Gene. Molecular & General Genetics **230**, 329-331.

Gupta, S., Gallavotti, A., Stryker, G.A., Schmidt, R.J., and Lal, S.K. (2005). A novel class of Helitron-related transposable elements in maize contain portions of multiple pseudogenes. Plant Molecular Biology **57**, 115-127.

Haberer, G., Young, S., Bharti, A.K., Gundlach, H., Raymond, C., Fuks, G., Butler, E., Wing, R.A., Rounsley, S., Birren, B.*, et al.* (2005). Structure and architecture of the maize genome. Plant Physiol **139**, 1612-1624.

Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R.K., Lisch, D., Meyers, B.C., Shiu, S.H., and Jiang, N. (2009). The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. Plant Cell **21**, 25-38.

Hartmann, U., Hohmann, S., Nettesheim, K., Wisman, E., Saedler, H., and Huijser, P. (2000). Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. Plant J. **21**, 351-360.

Hartwig, T., Chuck, G.S., Fujioka, S., Klempien, A., Weizbauer, R., Potluri, D.P., Choe, S., Johal, G.S., and Schulz, B. (2011). Brassinosteroid control of sex determination in maize. Proc. Natl. Acad. Sci. USA **108**, 19814-19819.

He, C., Munster, T., and Saedler, H. (2004). On the origin of floral morphological novelties. FEBS Lett. **567**, 147-151.

Heinlein, M. (1996). Excision patterns of Activator (Ac) and Dissociation (Ds) elements in Zea mays L.: implications for the regulation of transposition. Genetics **144**, 1851-1869.

Hershberger, R.J., Warren, C.A., and Walbot, V. (1991). Mutator activity in maize correlates with the presence and expression of the Mu transposable element Mu9. Proc Natl Acad Sci U S A **88**, 10198-10202.

Hoen, D.R., Park, K.C., Elrouby, N., Yu, Z.H., Mohabir, N., Cowan, R.K., and Bureau, T.E. (2006). Transposon-mediated expansion and diversification of a family of ULP-like genes. Molecular Biology and Evolution **23**, 1254-1268.

Houbaherin, N., Becker, D., Post, A., Larondelle, Y., and Starlinger, P. (1990). Excision of a Ds-Like Maize Transposable Element (Ac-Delta) in a Transient Assay in Petunia Is Enhanced by a Truncated Coding Region of the Transposable Element Ac. Molecular & General Genetics **224**, 17-23.

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.

Hudson, M.E., Lisch, D.R., and Quail, P.H. (2003). The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant Journal **34**, 453-471.

Jeddeloh, J.A., Bender, J., and Richards, E.J. (1998). The DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis. Gene Dev **12**, 1714-1725.

Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., and Wessler, S.R. (2004). Pack-MULE transposable elements mediate gene evolution in plants. Nature **431**, 569-573.

Jiang, N., Ferguson, A.A., Slotkin, R.K., and Lisch, D. (2011). Pack-Mutator-like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition. P Natl Acad Sci USA **108**, 1537-1542.

Jones, J.D., Carland, F.M., Maliga, P., and Dooner, H.K. (1989). Visual detection of transposition of the maize element activator (ac) in tobacco seedlings. Science **244**, 204-207.

Jurka, J., Kapitonov, V.V., Kohany, O., and Jurka, M.V. (2007). Repetitive sequences in complex genomes: structure and evolution. Annu Rev Genomics Hum Genet **8**, 241-259.

Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. P Natl Acad Sci USA **98**, 8714-8719.

Kapitonov, V.V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. P Natl Acad Sci USA **103**, 4540-4545.

Kapitonov, V.V., and Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet **23**, 521-529.

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

Kidwell, M.G., and Lisch, D.R. (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution **55**, 1-24.

Kleckner, N. (1990). Regulating tn10 and is10 transposition. Genetics **124**, 449-454.

Kolkman, J.M., Conrad, L.J., Farmer, P.R., Hardeman, K., Ahern, K.R., Lewis, P.E., Sawers, R.J.H., 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.

Konieczny, A., and Ausubel, F.M. (1993). A Procedure for Mapping Arabidopsis Mutations Using Codominant Ecotype-Specific Pcr-Based Markers. Plant J. **4**, 403-410.

Kunze, R., and Starlinger, P. (1989). The putative transposase of transposable element Ac from Zea mays L. interacts with subterminal sequences of Ac. EMBO J **8**, 3177-3185.

Kunze, R., Stochaj, U., Laufs, J., and Starlinger, P. (1987). Transcription of transposable element Activator (Ac) of Zea mays L. EMBO J **6**, 1555-1563.

Kunze, R., and Weil, C. (2002). Mobile DNA (Washington, D.C.: ASM Press).

Lai, J.S., Li, Y.B., Messing, J., and Dooner, H.K. (2005). Gene movement by Helitron transposons contributes to the haplotype variability of maize. P Natl Acad Sci USA **102**, 9068-9073.

Lal, S.K., Giroux, M.J., Brendel, V., Vallejos, C.E., and Hannah, L.C. (2003). The maize genome contains a Helitron insertion. Plant Cell **15**, 381-391.

Langdale, J.A., Irish, E.E., and Nelson, T. (1994). Action of the Tunicate Locus on Maize Floral Development. Dev. Genet. **15**, 176-187.

Le Roux, L.G., and Kellogg, E.A. (1999). Floral development and the formation of unisexual spikelets in the Andropogoneae (Poaceae). Am. J. Bot. **86**, 354-366.

Leiva, J., Dante, R., and Holding, D.R. (2002). RNA extraction from maize endosperm using SDS-TRIZOL protocol.

Lim, J.K., and Simmons, M.J. (1994). Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster. Bioessays **16**, 269-275.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., *et al.* (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430**, 471-476.

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

Lisch, D. (2005). Pack-MULEs: theft on a massive scale. Bioessays **27**, 353-355.

Lisch, D., Carey, C.C., Dorweiler, J.E., and Chandler, V.L. (2002). A mutation that prevents paramutation in maize also reverses Mutator transposon methylation and silencing. P Natl Acad Sci USA **99**, 6130-6135.

Lisch, D., Chomet, P., and Freeling, M. (1995). Genetic characterization of the Mutator system in maize: behavior and regulation of Mu transposons in a minimal line. Genetics **139**, 1777-1796.

Liu, S.Z., Yeh, C.T., Ji, T.M., Ying, K., Wu, H.Y., Tang, H.M., Fu, Y., Nettleton, D., and Schnable, P.S. (2009). Mu Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome. Plos Genet **5**.

Mangelsdorf, P.C. (1947). The origin and evolution of maize. Adv. Genet. **1**, 161-207.

Mangelsdorf, P.C. (1974). Corn: Its Origin, Evolution and Improvement. (Cambridge: Harvard University Press).

Mangelsdorf, P.C. (1984). The origin of maize. Science **225**, 1094.

Mangelsdorf, P.C., and Galinat, W.C. (1964). The Tunicate Locus in Maize Dissected and Reconstituted. Proc. Natl. Acad. Sci. USA **51**, 147-150.

Mangelsdorf, P.C., Macneish, R.S., and Galinat, W.C. (1964). Domestication of Corn. Science **143**, 538-545.

Martienssen, R., Barkan, A., Taylor, W.C., and Freeling, M. (1990). Somatically heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. Genes Dev **4**, 331-343.

Martienssen, R., and Baron, A. (1994). Coordinate suppression of mutations caused by Robertson's mutator transposons in maize. Genetics **136**, 1157-1170.

Martienssen, R.A., Rabinowicz, P.D., O'Shaughnessy, A., and McCombie, W.R. (2004). Sequencing the maize genome. Curr Opin Plant Biol **7**, 102-107.

Masson, P., Surosky, R., Kingsbury, J.A., and Fedoroff, N.V. (1987). Genetic and molecular analysis of the Spm-dependent a-m2 alleles of the maize a locus. Genetics **117**, 117-137.

Matsumoto, T., Wu, J.Z., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., Mizuno, H., Yamamoto, K., Antonio, B.A., Baba, T.*, et al.* (2005). The map-based sequence of the rice genome. Nature **436**, 793-800.

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

May, B.P., and Martienssen, R.A. (2003). Transposon mutagenesis in the study of plant development. Crit Rev Plant Sci **22**, 1-35.

May, E.W., and Craig, N.L. (1996). Switching from cut-and-paste to replicative Tn7 transposition. Science **272**, 401-404.

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

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A **36**, 344-355.

McClintock, B. (1954). Mutations in maize and chromosomal observations in Neurospora., Vol 53 (Carnegie Institute of Washington Year Book).

McClintock, B. (1956). Controlling Element and the Gene. Cold Spring Harbor Symposia on Quantitative Biology **21**, 197-216.

McClintock, B. (1961). Further studies on the suppressor-mutator system of control of gene action in maize, Vol 60 (Carnegie Institute of Washington Year Book).

McClintock, B. (1965). The control of gene action in maize. Brookhaven Symp. Biol **18**, 162-184.

McClintock, B. (1984). The significances of responses of the genome to challenge. Science **226**, 792–801.

Menssen, A., Hohmann, S., Martin, W., Schnable, P.S., Peterson, P.A., Saedler, H., and Gierl, A. (1990). The En/Spm Transposable Element of Zea-Mays Contains Splice Sites at the Termini Generating a Novel Intron from a Dspm Element in the A2 Gene. Embo Journal **9**, 3051-3057.

Messing, J., Bharti, A.K., Karlowski, W.M., Gundlach, H., Kim, H.R., Yu, Y., Wei, F.S., Fuks, G., Soderlund, C.A., Mayer, K.F.X., *et al.* (2004). Sequence composition and genome organization of maize. P Natl Acad Sci USA **101**, 14349-14354.

Meyers, B.C., Tingey, S.V., and Morgante, M. (2001). Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res **11**, 1660-1676.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature **411**, 212-214.

Mohanty, A., Luo, A., DeBlasio, S., Ling, X., Yang, Y., Tuthill, D.E., Williams, K.E., Hill, D., Zadrozny, T., Chan, A., *et al.* (2009). Advancing cell biology and functional genomics in maize using fluorescent protein-tagged lines. Plant Physiol. **149**, 601-605.

Moreno, M.A., Chen, J., Greenblatt, I., and Dellaporta, S.L. (1992). Reconstitutional Mutagenesis of the Maize P-Gene by Short-Range Ac Transpositions. Genetics **131**, 939-956.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., and Rafalski, A. (2005). Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nature Genetics **37**, 997-1002.

Munster, T., Deleu, W., Wingen, L.U., Ouzunova, M., Cacharron, J., Faigl, W., Wert, S., Kim, J.T.T., Saedler, H., and Theissen, G. (2002). Maize MADS-box genes galore. Maydica **47**, 287-301.

Munster, T., Wingen, L., Faigl, W., Deleu, W., Ahlbory, D., Sommer, H., Saedler, H., and Theissen, G. (2004). Pod corn (Tunicate maize) is caused by the ectopic expression of a vegetative MADS-box gene in the inflorescence of maize. Genbank AJ850299-AJ850303.

Nassif, N., Penney, J., Pal, S., Engels, W.R., and Gloor, G.B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol **14**, 1613-1625.

Ng, M., and Yanofsky, M.F. (2001). Function and evolution of the plant MADS-box gene family. Nat .Rev. Genet. **2**, 186-195.

Nickerson, N.H., and Dale, E.E. (1955). Tassel Modifications in Zea Mays. Ann. Mo. Bot. Gard. **42**, 195-211.

Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., De Paoli, E., Accerbi, M., Arteaga-Vazquez, M., Sidorenko, L., Jeong, D.H., Yen, Y., *et al.* (2008). Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. Proc Natl Acad Sci U S A **105**, 14958-14963.

O'Reilly, C., Shepherd, N.S., Pereira, A., Schwarz-Sommer, Z., Bertram, I., Robertson, D.S., Peterson, P.A., and Saedler, H. (1985). Molecular cloning of the a1 locus of Zea mays using the transposable elements En and Mu1. EMBO J **4**, 877-882.

Palmer, L.E., Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V.S., Nascimento, L.U., Dike, S., de la Bastide, M., Martienssen, R.A., and McCombie, W.R. (2003). Maize genome sequencing by methylation filtration. Science **302**, 2115-2117.

Parinov, S., and Sundaresan, V. (2000). Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotech **11**, 157-161.

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., Ainouche, M., Chalhoub, B., and Grandbastien, M.A. (2010). Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol **186**, 37-45.

Peterson, P.A. (1953). A mutable pale green locus in maize. Genetics **38**, 682-683.

Plasterk, R.H.A. (1991). The Origin of Footprints of the Tc1 Transposon of Caenorhabditis-Elegans. Embo Journal **10**, 1919-1925.

Pollock, R., and Treisman, R. (1991). Human Srf-Related Proteins - DNA-Binding Properties and Potential Regulatory Targets. Gene Dev **5**, 2327-2341.

Pritham, E.J., Putliwala, T., and Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene **390**, 3-17.

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

Raboy, V., Kim, H.Y., Schiefelbein, J.W., and Nelson, O.E. (1989). Deletions in a Dspm Insert in a Maize Bronze-1 Allele Alter Rna Processing and Gene-Expression. Genetics **122**, 695-703.

Raizada, M.N., Nan, G.L., and Walbot, V. (2001). Somatic and germinal mobility of the RescueMu transposon in transgenic maize. Plant Cell **13**, 1587-1608.

Roccaro, M., Li, Y., Sommer, H., and Saedler, H. (2007). ROSINA (RSI) is part of a CACTA transposable element, TamRSI, and links flower development to transposon activity. Molecular Genetics and Genomics **278**, 243-254.

Rowold, D.J., and Herrera, R.J. (2000). Alu elements and the human genome. Genetica **108**, 57-72.

Rudenko, G.N., and Walbot, V. (2001). Expression and post-transcriptional regulation of maize transposable element MuDR and its derivatives. Plant Cell **13**, 553-570.

Saint-Hilaire, A.d. (1829). Lettre sur une variete remarquable de Mais du Bresil. Ann. Sci. Nat. **16**, 143-145.

Sanmiguel, P., and Bennetzen, J.L. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot-London **82**, 37-44.

Sasaki, H., and Watanabe, T. (2008). Endogenous siRNAs derived from naturally formed dsRNAs regulate coding-transcripts and retrotransposons in mouse oocytes. Genes Genet Syst **83**, 532-532.

Satoh-Nagasawa, N., Nagasawa, N., Malcomber, S., Sakai, H., and Jackson, D. (2006). A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature **441**, 227-230.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., *et al.* (2009a). The B73 maize genome: complexity, diversity, and dynamics. Science **326**, 1112-1115.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F.S., Pasternak, S., Liang, C.Z., Zhang, J.W., Fulton, L., Graves, T.A., *et al.* (2009b). The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science **326**, 1112-1115.

Schwartz, D. (1989). Gene-controlled cytosine demethylation in the promoter region of the Ac transposable element in maize. Proc Natl Acad Sci U S A **86**, 2789-2793.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N., and Kaeppler, S.M. (2011). Genome-wide atlas of transcription during maize development. Plant J. **66**, 553-563.

Sentoku, N., Kato, H., Kitano, H., and Imai, R. (2005). OsMADS22, an STMADS11-like MADS-box gene of rice, is expressed in non-vegetative tissues and its ectopic expression induces spikelet meristem indeterminacy. Mol Genet Genomics **273**, 1-9.

Settles, A.M. (2005). Maize Community Resources for Forward and Reverse Genetics. Maydica **50**, 405-414.

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.*, et al.* (2007). Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. Bmc Genomics **8**.

Shapiro, J., and MacHattie, L. (1979). Integration and excision of prophage lambda mediated by the IS 1 element. Cold Spring Harb Symp Quant Biol **43**, 1135-1142.

Shapiro, J.A. (1969). Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. J Mol Biol **40**, 93-105.

Shore, P., and Sharrocks, A.D. (1995). The Mads-Box Family of Transcription Factors. Eur J Biochem **229**, 1-13.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., and Lukyanov, S.A. (1995). An Improved Pcr Method for Walking in Uncloned Genomic DNA. Nucleic Acids Res **23**, 1087-1088.

Sijen, T., and Plasterk, R.H.A. (2003). Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature **426**, 310-314.

Singer, T., Yordan, C., and Martienssen, R.A. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev **15**, 591-602.

Song, R.T., and Messing, J. (2003). Gene expression of a gene family in maize based on noncollinear haplotypes. P Natl Acad Sci USA **100**, 9055-9060.

Swinburne, J., Balcells, L., Scofield, S.R., Jones, J.D., and Coupland, G. (1992). Elevated levels of Activator transposase mRNA are associated with high frequencies of Dissociation excision in Arabidopsis. Plant Cell **4**, 583-595.

Szustakowski, J., and Consor, I.H.G.S. (2001). Initial sequencing and analysis of the human genome (vol 409, pg 860, 2001). Nature **411**, 720-720.

Talbert, L.E., and Chandler, V.L. (1988). Characterization of a highly conserved sequence related to mutator transposable elements in maize. Mol Biol Evol **5**, 519-529.

Trevaskis, B., Tadege, M., Hemming, M.N., Peacock, W.J., Dennis, E.S., and Sheldon, C. (2007). Short vegetative phase-like MADS-box genes inhibit floral meristem identity in barley. Plant Physiol. **143**, 225-235.

Turcotte, K., Srinivasan, S., and Bureau, T. (2001). Survey of transposable elements from rice genomic sequences. Plant Journal **25**, 169-179.

Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev **20**, 759-771.

Volff, J.N. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays **28**, 913-922.

Vollbrecht, E., Duvick, J., Schares, J.P., Ahern, K.R., Deewatthanawong, P., Xu, L., Conrad, L.J., Kikuchi, K., Kubinec, T.A., Hall, B.D.*, et al.* (2010). Genome-Wide Distribution of Transposed Dissociation Elements in Maize. Plant Cell **22**, 1667-1685.

Vollbrecht, E., Springer, P.S., Goh, L., Buckler, E.S.t., and Martienssen, R. (2005). Architecture of floral branch systems in maize and related grasses. Nature **436**, 1119-1126.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). Arabidopsis-Thaliana DNA Methylation Mutants. Science **260**, 1926-1928.

Walbot, V. (2000). Saturation mutagenesis using maize transposons. Curr Opin Plant Biol **3**, 103-107.

Walbot, V., and Rudenko, G.N. (2002). MuDR/Mu Tranasposable Elements of Maize. In Mobile DNA II, N.L. Craig, ed. (ASM Press, Wasington, D.C.), pp. 533-564.

Walbot, V., and Warren, C. (1988). Regulation of Mu element copy number in maize lines with an active or inactive Mutator transposable element system. Mol Gen Genet **211**, 27-34.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T.*, et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature **453**, 539-U539.

Wei, F., Stein, J.C., Liang, C., Zhang, J., Fulton, R.S., Baucom, R.S., De Paoli, E., Zhou, S., Yang, L., Han, Y., *et al.* (2009). Detailed analysis of a contiguous 22-Mb region of the maize genome. PLoS Genet. **5**, e1000728.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.G., and Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. Cell **149**, 101-112.

Weil, C.F., and Wessler, S.R. (1993). Molecular Evidence That Chromosome Breakage by Ds Elements Is Caused by Aberrant Transposition. Plant Cell **5**, 515-522.

Wessler, S.R. (2006). Transposable elements and the evolution of eukaryotic genomes. Proc Natl Acad Sci U S A **103**, 17600-17601.

Whipple, C.J., Ciceri, P., Padilla, C.M., Ambrose, B.A., Bandong, S.L., and Schmidt, R.J. (2004). Conservation of B-class floral homeotic gene function between maize and Arabidopsis. Development **131**, 6083-6091.

Wicker, T., Guyot, R., Yahiaoui, N., and Keller, B. (2003). CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol **132**, 52-63.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., *et al.* (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet **8**, 973-982.

Wingen, L.U., Munster, T., Faigl, W., Deleu, W., Sommer, H., Saedler, H., and Theissen, G. (2012). Molecular genetic basis of pod corn (Tunicate maize). Proc. Natl. Acad. Sci. USA **109**, 7115-7120.

Wolfgruber, T.K., Sharma, A., Schneider, K.L., Albert, P.S., Koo, D.H., Shi, J.H., Gao, Z., Han, F.P., Lee, H., Xu, R.H., *et al.* (2009). Maize Centromere Structure and Evolution: Sequence Analysis of Centromeres 2 and 5 Reveals Dynamic Loci Shaped Primarily by Retrotransposons. Plos Genet **5**.

Woodhouse, M.R., Freeling, M., and Lisch, D. (2006). The mop1 (mediator of paramutation1) mutant progressively reactivates one of the two genes encoded by the MuDR transposon in maize. Genetics **172**, 579-592.

Xu, J.H., and Messing, J. (2006). Maize haplotype with a helitron-amplified cytidine deaminase gene copy. BMC Genet **7**, 52.

Xu, Z.N., Yan, X.H., Maurais, S., Fu, H.H., O'Brien, D.G., Mottinger, J., and Dooner, H.K. (2004). Jittery, a mutator distant relative with a paradoxical mobile behavior: Excision without reinsertion. Plant Cell **16**, 1105-1114.

Yoder, J.I. (1990). Rapid proliferation of the maize transposable element Activator in transgenic tomato. Plant Cell **2**, 723-730.

Yu, H., Ito, T., Wellmer, F., and Meyerowitz, E.M. (2004). Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. Nat Genet **36**, 157-161.

Yu, H., Xu, Y., Tan, E.L., and Kumar, P.P. (2002). AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. Proc. Natl. Acad. Sci. USA **99**, 16336-16341.

Yu, Z., Wright, S.I., and Bureau, T.E. (2000). Mutator-like elements in Arabidopsis thaliana. Structure, diversity and evolution. Genetics **156**, 2019-2031.

Zabala, G., and Vodkin, L.O. (2005). The wp mutation of Glycine max carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell **17**, 2619-2632.

Zhang, J.B., and Peterson, T. (2004). Transposition of reversed Ac element ends generates chromosome rearrangements in maize. Genetics **167**, 1929-1937.

Zhu, Q.H., and Helliwell, C.A. (2011). Regulation of flowering time and floral patterning by miR172. J. Exp. Bot. **62**, 487-495.

Appendix

I. Statement of Work (SOW)

Title

Investigation of the origin of pod corn in North America

Abstract

Pod corn, once considered as ancestral to cultivated corn, has a striking phenotype enclosing each kernel with chaff. It has been preserved in Central and South American cultures due to its spiritual and traditional properties. Genetic research in the 1920s revealed that pod corn is a well-known classic maize mutant *Tunicate1* (*Tu1*). Our recent work[1] has revealed that *Tu1* is explained by a chromosomal (DNA sequence) rearrangement at the *Tu1* locus that alters the expression pattern of the *Tu1* gene involved in determining the identity of flowers in cobs and tassels. Recently, we obtained 2 pod corn herbarium specimens from the Missouri Botanical Garden, which were collected in Peru and Mexico. We found molecular evidence of the rearranged *Tu1* in these samples. Now we are interested in tracking ancient samples of pod corn, collected by the Museum of Northern Arizona at Flagstaff from the Betatakin pueblo and dating from the 14th century. In this project, we will investigate genetic modifications of pod corn to study introduction and domestication of pod corn in North American, which has been the subject of speculation since the specimens were found in the 1940s.

Keywords –biology, evolution, pod corn, half tunicate, domestication

Statement of issue

Pod corn was once regarded as the primitive ancestor of modern cultivated corn due to its characteristic of enclosing the kernel in glumes and was preserved by pre-Columbian

peoples as a remedy for respiratory conditions known as "the air"[2] and for other magical properties. The *Tunicate1* (*Tu1*) mutant of maize is a naturally-occurring dominant mutation, resulting in a striking pod corn phenotype. In a dosage-dependent manner, *Tu1* results in glume elongation, sex indeterminacy, and branching defects in tassel and ear. Mangelsdorf and his collegues[3,4] provided a genetic evidence of a dosage effect on *Tu1* phenotype by generating a weak form of pod corn that is called "half tunicate" (Fig. 1). We found that the *Tu1* locus consists of the maize MADS-box gene, *Zmm19*, whose regulatory region is fused by chromosomal inversion to another gene that is located about 2 million bases away on the same chromosome. Interestingly, our genetic mapping analysis[1] suggested that the insertion of a transposable element may be the cause of the chromosomal inversion. This inversion may have prevented fine-mapping of this candidate gene in previous studies[5,6] because rearrangement prohibits recombination, and likely accompanied the origin of this ancient gene. Our genetic and molecular data[1] indicated that misregulation of *Tu1* gives rise to pleiotropic pod corn phenotypes.

We showed that *Tu1* transgenic maize plants exactly resemble half-tunicate pod corn (Fig. 2B and 2C). We also observed an additive effect of phenotypic trait from the combination of *Tu1-A* and *Tu1-B* (Fig. 2D). TU1 proteins fused to red and yellow fluorescent proteins (RFP, YFP) are co-localized to the nucleus in young tassels and cobs, where it controls flower development. Specifically, YFP and RFP-fused TU1 proteins are expressed in discrete cup-shaped domains at the base of spikelet pair meristems in young ear primordia.

Figure 1 Phenotypes. A domesticated form of modern corn showing the cob with very short glumes (left) and a weak form of pod corn showing the cob with elongated glumes (right).

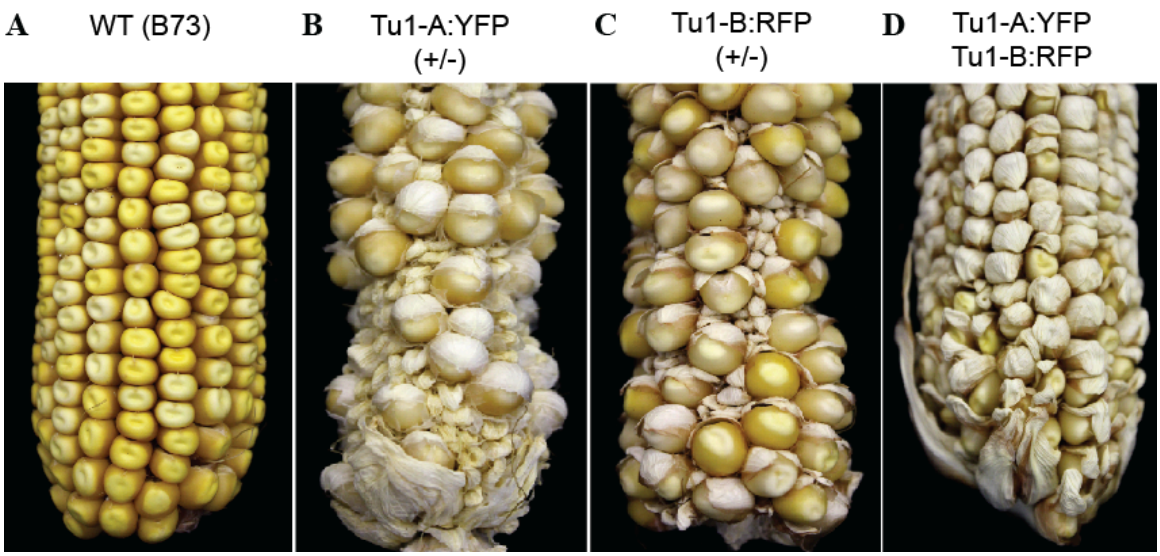


Figure 2 Phenotypes of Tu1 transgenic mutant. A, Maize fertilized wild-type ear. **B,** *Tu1* transgenic ear with *Tu1-A* allele displaying half-tunicate phenotype with elongated glumes. **C,** *Tu1* transgenic ear with *Tu1-B* allele also displaying half-tunicate phenotype. **D,** *Tu1* transgenic ear with both *Tu1-A* and *Tu1-B* allele showing additive effect to fully enclose kernels.

Recently, we obtained 2 pod corn herbarium specimens from the Missouri Botanical Garden, which were collected in Peru and Mexico (Fig 3). Since our previous study revealed that a novel *mutator-like* DNA transposon is located at the 5' upstream region of maize MADS-box gene *Zmm19*[1], we could design 2 pairs of primers specific to *Tu1* allele, which recognize left and right junction of the novel DNA transposon in *Zmm19* upstream region, respectively. After extracting DNA from these 100 year-old herbarium specimens, we applied our DNA test by performing Polymerase Chain Reaction (PCR) with 2 primer sets, resulting in successful amplification of *Tu1* allele from pod corn but not from normal specimens (Fig 4).

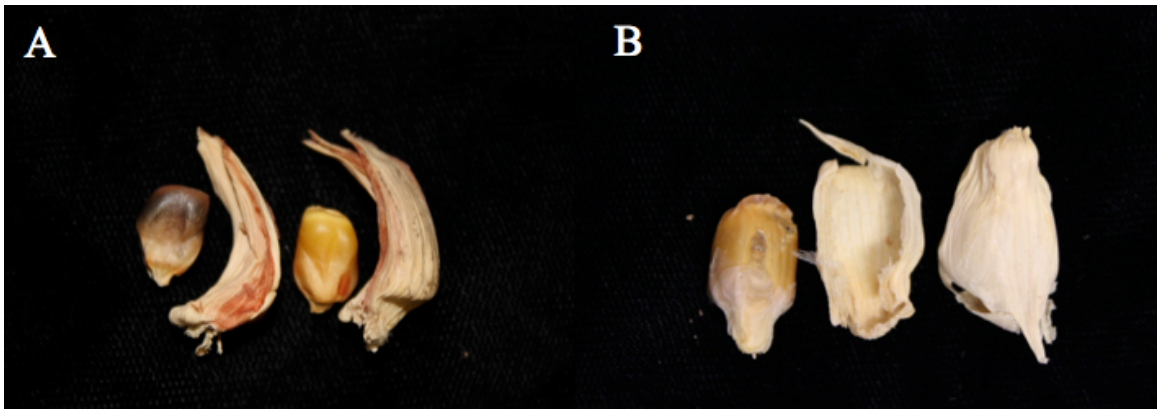


Figure 3 Phenotypes of pod corn herbarium specimens. A, Pod corn specimen from Peru. **B,** Pod corn specimen from Mexico.

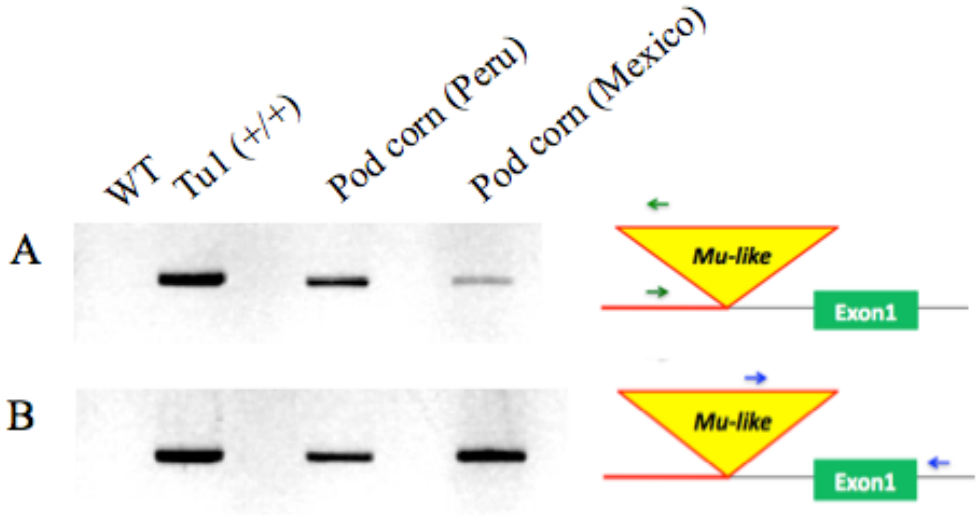


Figure 4 PCR analysis of pod corn herbarium specimens. A, Left junction of a novel *Mu-like* element is present in both specimens. A reverse triangle displays the position of a novel *Mu-like* element with green arrows showing the position of primers. **B,** Right junction of a novel *Mu-like* element is also present in both specimens. Blue arrows present the position of primers.

A few archaeological specimens of pod corn have been reported so far in North America. One of them is from Richards' Caves near Montezuma's Castle, Arizona[7] and the other

from Betatakin, a large pueblo ruin in Tsegi Canyon, northeastern Arizona[2]. Cutler (1944)[2] suggested that the ancient Southwestern hunchbacked flute player sometimes called "Kokopelli" might be Callahuavo medicine men who traveled through an ancient trade route to South, Central, and even North America carrying in their packs common remedies including pod corn. If this was true, it would be plausible for pod corn to have spread from its original home to all parts of the North America. Ears of pod corn from Betatakin are preserved in the Museum of Northern Arizona. It would be of great interest to determine if these specimens have the same gene to track the ancestor of *Zea mays* var. *tunicata*.

Scope of study

We will request ancient ear cobs of pod corn to the Museum of Northern Arizona. Using the samples, we will extract DNA, amplify the plausible junctions of chromosomal inversion, and sequence the amplified DNA products. A series of experimental activities will take place in the Cold Spring Harbor Laboratory. We may also attempt to use "next generation sequencing" technology to determine the DNA sequence of a larger part of the pod corn genome. This will help determine the degree of domestication and relatedness to modern North, Central and South American corn landraces.

Objective/Hypothesis to be tested

1. Extracting DNA from ancient samples of pod corn
   - whether ancient DNAs can be purely extracted from the ancient specimens
2. Identifying the presence of a transposon insertion in regulatory region of *Tu1* gene
   - whether ancient pod corns have the same transposon that is present in descendent pod corns
3. Investigating the presence of the *Tu1* locus
   - whether the ancient pod corns passed a chromosomal inversion down to their descendents

4. Mapping ancient pod corn DNA sequences against Maize whole genome sequence data from modern inbreds and landraces
   - whether genetic signatures of ancient pod corn can be found in their descendents among existing landraces , genome-wide

Methods

Between 0.2 g and 1 g of ancient cob specimens[8] is powdered and rotated 15h at 37°C in 4ml extraction buffer (Tris-HCl, pH8.0/10mM NaCl/10mM dithiothreitol containing proteinaseK at 0.4 mg/ml and 1% sodium dodecyl sulfate)[9] containing 10mM EDTA and 2.5mM PTB[10]. The sample is extracted twice with phenol(pH8.0;1:1,vol/vol) and once with chloroform/isoamyl alcohol (24:1, vol/vol). The DNA is retrieved by binding to silica[11,12]. PCR is performed[13] using the following primer pairs and annealing temperatures; Tu1 5' F1, 5'-AGA GTC GTA CTA CTC TGC AAG TGC -3' and Tu1 5' R1, 5'-TGC AGT TAT GTG TTG GTG TTA TGG -3' at 60°C. These primers amplify a 823bp segment in the 5' upstream region of *Zmm19* gene with the left junction of a novel *Mu-like* element; Tu1 5' F3, 5'-TTA GAA GTG GTG GAT GCA TGA AGA -3' and Tu1 5' R3, 5'-AGG GAA GGG AAC TGC ATA CTG TGT –3' at 60°C. These primers amplify a 1218bp segment between the 5' upstream region of *Zmm19* gene and the right junction of a novel *Mu-like* element. They cannot amplify 5' upstream region of wild-type *Zmm19* allele.

Each gene fragment is amplified independently at least twice from each sample and cloned into TOPO-TA vector for sequencing. At least 20 clones are sequenced from each product. Cobs are considered to be heterozygous if two independent PCR-products show the same two alleles. Cobs that only show one sequence in both amplifications are considered to be homozygous although heterozygosity cannot be rigorously excluded[14].

Analysis

Gramene (www.gramene.org) is a combined resource for genetic, genomic, and comparative genomic data for major crop species, such as maize (*Zea mays*), sorghum (*Sorghum*

bicolar), barley (*Hordeum vulgare*), wheat (*Triticum aestivum*), oat (*Avena sativa*), and potentially teosinte (maize ancestor). We will utilize this comparative genome mapping database for grasses to study genetic diversity of ancient pod corn DNA among other grass genome sequences. These comparative analyses will provide us relevance of ancient pod corn to modern corn landraces and other grasses in an evolutionary aspect, as well as genetic/genomic variations of ancient pod corn on domestication and specification of modern corn landraces. Large-scale genome comparisons along with regional *Tu1* DNA sequence homology search will open up the unknown historical position of the ancient pod corn along the timeline of maize evolution.

Literature cited

1. **Han, J-.J., Jackson, D., Martienssen, R.** (2012) Pod corn is caused by rearrangement at the *Tunicate1* locus. Plant Cell **24**, 2733-2744
2. **Cutler, H.C.** (1944)**.** Medicine men and the preservation of a relict gene in Maize. Journal of Heredity. **35**, 291-294.
3. **Mangelsdorf, P.C.** (1947). The origin and evolution of maize. Adv. Genet. **1**, 161-207.
4. **Mangelsdorf, P.C., and Galinat, W.C.** (1964). The Tunicate Locus in Maize Dissected and Reconstituted. Proc. Natl. Acad. Sci. USA **51**, 147-150.
5. **Wingen, L.U., Munster, T., Faigl, W., Deleu, W., Sommer, H., Saedler, H., and Theissen, G.** (2012). Molecular genetic basis of pod corn (Tunicate maize). Proc. Natl. Acad. Sci. USA **109**, 7115-7120.
6. **Munster, T., Deleu, W., Wingen, L.U., Ouzunova, M., Cacharron, J., Faigl, W., Wert, S., Kim, J.T.T., Saedler, H., and Theissen, G.** (2002). Maize MADS-box genes galore. Maydica **47**, 287-301.
7. **Galinat, W.C., Mangelsdorf, P.C., and Pierson, L.** (1956). Estimates of teosinte introgression in archaeological maize, Botanical Museum Leaflets of Harvard University **17**(4), 101-124.

8. **Jaenicke-Despres, V., Buckler E.S., Smith, B.D., Gilbert M.T.P., Cooper, A., Doebley, J., Pääbo, S.** (2003). Early Allelic Selection in Maize as revealed by ancienct DNA. Science **302**, 1206-1208.
9. **Goloubinoff, P., Pääbo, S., and Wilson, A.C.** (1993). Evolution of maize inferred from sequence diversitz of an Adh2 gene segment from archaeological specimens. Proc. Natl. Acad. Sci. USA **90,** 1997-2001.
10. **Poinar H.N., Schwarz, C., Qi, J., Shapiro, B., MacPhee R.D.E., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho L.P., Auch, A., Rampp, M., Miler, W., Schuster, S.C.** (2006), Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. Science **281**, 402-406.
11. **Höss, M., Pääbo, S.** (1993). DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic acid research **21**, 3913-3914.
12. **Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Wertheim-van Dillen, P.M.E. and van der Noordaa, J.** (1990). Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. **28**, 495-503.
13. **Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A., Pääbo S.** (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic acid research **29**, 4793-4799.
14. **Morin, P. A., Charmbers, K. E., Boesch, C., Vigilant, L.** (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Molecular Ecology **10**, 1835–1844.