A COMPARATIVE GENOMIC STUDY OF DOUBLE-STRAND BREAK REPAIR IN MAIZE AND RICE

by

FANG LU

(Under the Direction of Jeffrey L. Bennetzen)

ABSTRACT

Double-strand break (DSB) repair is essential for cell survival and for the maintenance of genome integrity. In this study, we utilized I-*Sce* I, an endonuclease from yeast with an 18-bp recognition site to introduce DSB in maize and rice cells. We then employed PCR to characterize the inaccurate repair events that removed the I-*Sce* I recognition site. Of 82 and 84 inaccurate repair events characterized in maize and rice, more than 72% inaccurate repairs were associated with deletions. Of the deletion repairs, small deletions (1-9 nucleotides) happened more frequently than large deletions. The sequences flanking these deletions and insertions have the hallmarks of illegitimate recombination. These results are compatible with models suggesting that inaccurate repair of DSBs is the major factor responsible for the rapid removal of unselected DNA from higher plant genomes. The sizes of inaccurate repairs are not necessarily associated with genome size variation in monocot species.

INDEX WORDS: Comparative Genomics, Double-Strand Breaks (DSBs), I-*Sce* I, Maize, Rice

A COMPARATIVE GENOMIC STUDY OF DOUBLE-STRAND BREAK REPAIR IN MAIZE AND RICE

by

FANG LU

Bachelor of Agricultural Science, Beijing Agricultural College, 1994

Master of Agricultural Science, Chinese Academy of Agricultural Sciences, 1997

Master of Science, University of Georgia, 2008

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2009

A COMPARATIVE GENOMIC STUDY OF DOUBLE-STAND BREAK REPAIR IN MAIZE AN RICE

by

FANG LU

Major Professor: Jeffery L. Bennetzen

Committee: Kelly Dawe
Michael McEachern

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2009

DEDICATION

To my husband Shunxue Tang for his accompany, accommodation, patience, and unlimited support;

To my mother Ruqin Jiang and my father Deyang Lu for their unconditional love, always encouraging me to work hard and accomplish my goals.

ACKNOWLEDGEMENTS

I would thank my advisor, Dr. Jeffery Bennetzen, for his continuing support to me to finish my studies at UGA; for his guidance throughout the research project, especially when I was in troubles during my research work; for his insightful knowledge, creative thought, broad experience, meticulous attitude and intelligent talks in research, for his tireless effort in sustainable agriculture development, as well as for his good personality, which make everybody working happily under his direction. This beneficial and cherishing research experience opened a door for me in molecular biology and genomic research.

I would like to extend my sincere thanks to my committee members, Dr. Kelly Dawe, for his insights and knowledge in maize clonal analysis and plant genetics, precious encouragement and unselfish help; to Dr. Michael McEachern, for his broad knowledge in DNA recombination, and timely support always.

Special acknowledgement goes to Dr. Wayne Parrott, for his broad knowledge in plant transformation technology and his generosity of allowing me to access his people and facility. I would like to thank Dr. Rodney Mauricio, for his suggestions on my data analysis, his continual care on my study.

I would express my sincere thank to all the other members in Dr. Bennetzen's lab, for sharing with me their experience, and making me have such a good life in the lab. I would like to thank Dr. Ervin Nagy and Dr Regina Baucom for their discussions and technical guidance; Ryan Percifield for his technology and managing support; Ansuya Jogi for her assistance in sampling**;** and Dr. Shavanor Smith, who introduced the maize crossing technique to me.

Special thanks to Eunyoung Cho and Sameer Khanal who taught me the rice crossing technique. Without them, I cannot finish my research project smoothly.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Double-strand breaks (DSBs) are one of the most common chromosomal events in living cells. DSBs can happen spontaneously during DNA replication and meiosis, or accidentally under stressful conditions such as X-rays or other mutagens. During meiosis, DSBs are closely associated with chromosomal recombination and segregation. In late S phase and G2 phase, a sister chromatid can be used as the template to recover all of the information lost at the break site. While in the G1 phase of somatic cells, no sister chromatids are available as template during the repair process. This situation makes DSB repair a fascinating study.

The DSB repair mechanisms are highly conserved through all of the kingdoms because of their vital importance to an organisms' survival. From prokaryotes to eukaryotes, from yeast to *Drosophila*, from worms to mammals and plants, all organisms have highly efficient DSB repair systems to fix chromosomal breaks under varied conditions. In plants, some disease resistance genes have changed in coding potential through illegitimate recombination (WICKER *et al.* 2007). Furthermore, numerous human diseases are associated with defects in DSB repair, such as the Ataxia Telangiectasia-Like Disorder (ATLD) and Nijmegen Breakage Syndrome (NBS) for mammals. These mutants with defective DSB repair have been a very good research model to study DSB repair mechanisms.

It is believed that the DSB repair mechanisms are an important factor in plant genome size change. Polyploidy, retrotransposon proliferation, chromosomal segmental duplication and

unequal homologous recombination are the most popular explanations for genome size changes (BENNETZEN *et al.* 2005; VITTE and BENNETZEN 2006). Homologous recombination (HR), which required significant sequence homology, and illegitimate recombination, which requires little or no sequence homology, are two primary pathways for DSB repair. Besides their action in DNA rearrangement, unequal homologous recombination and illegitimate recombination are the known genome-reducing mechanisms that counterbalance the genome size increases caused primarily by transposon amplification and by polyploidization (BENNETZEN *et al.* 2005; DEVOS *et al.* 2002).

Current research on DSB repair not only focuses on biochemical mechanisms, but also sheds new light on the profound role of DSB repair in genome evolutions. The chromosomal recombination initiated by DSBs also could make DSB repair a powerful tool to generate targeted mutations in any eukaryote, a very useful tool for functional genomics studies.

In this project, I initiated the study of DSB repair in four monocot plant species with varied genome size, in order to elucidate the role of DSB repair on the evolution of genome size and structure, through a comparative genomic method. I characterized the inaccurate repair events at I-*Sce* I-induced DSB sites in maize and rice, while future studies in the Bennetzen lab will pursue this same research in pearl millet and sorghum. By comparing the inaccurate repair events across these monocots, and comparison between the monocots and the dicotyledonous plants such as *Arabidopsis* and tobacco, we will have a more complete picture of the effect of DSB repair on genome evolution. Specifically, this research focused on the following:

Objective 1: Identifying and characterizing the DSB repair events in the cereals maize, rice, sorghum and millet, four monocot grass species. By comparing the DSB repair events between the species with different genome size, we can determine whether there is any

correlation between inaccurate repair outcomes and a genome size reducing mechanism, if an inverse relationship is seen between the deletion frequency and/or size and genome size.

Objective 2: Investigating any possible positional effects on inaccurate repair by mapping the different I-*Sce* I-induced break sites in the maize and rice genomes, then characterizing the outcomes of inaccurate repair at the different genomic sites. Different genomic sequence around DSB site at different location (e.g., repeat enriched heterochromatin region, or gene rich sequence) might be associated with some distinct characteristics of DSB repair.

CHAPTER 2

BACKGROUND, RATIONALE, AND SIGNIFICANCE

DSB might be induced by many factors, such as ionizing radiation or mutagenic chemicals. More significantly, DSB is a natural intermediate during many important cellular processes, such as DNA repair, DNA replication, meiotic recombination, and the excision of transposable elements. Chromosomal single-strand breaks, which frequently happen as an obligate component of base excision repair, nucleotide excision repair, and mismatch repair, as well as during DNA replication, in lagging strand synthesis, can also lead to the formation of DSBs, initiating the cell cycle checkpoint, followed by the DSB repair pathways. Failure to repair DSBs in animal cells will cause a cell cycle block and apoptosis (BREE *et al.* 2004; ROOS *et al.* 2009). Due to the key role it plays in these basic physiological processes, DSB repair is essential for all organisms' survival.

2.1 Homologous Recombination *Vs.* Illegitimate Recombination At DSBs

The biochemistry of DSB repair is well conserved across all of the kingdoms. Most of our knowledge on DSB repair mechanism comes from *E. coli*, lower eukaryotes, and mammalian cells. Although DSB repair in plants has been studied for a long time, there are still many gaps in our understanding. This is primarily due to the relatively few studies that have been undertaken.

Generally, DNA DSBs are repaired through homologous recombination (HR) and non-homologous end-joining (NHEJ). HR is believed to require a long tract of sequence homology (>30 nucleotides), while NHEJ does not. Both pathways can repair the chromosomal breaks faithfully or unfaithfully that generate chromosomal rearrangement. All the recombination events that do not require recombinase (RecA in *E. coli*, and Rad51, Dmc1in eukaryotes) and long sequence homology are called illegitimate recombination. It includes NHEJ and slippage strand replication (SSR). Inaccurate illegitimate recombination may lead to chromosomal rearrangement, such as deletions or insertions.

Homologous recombination

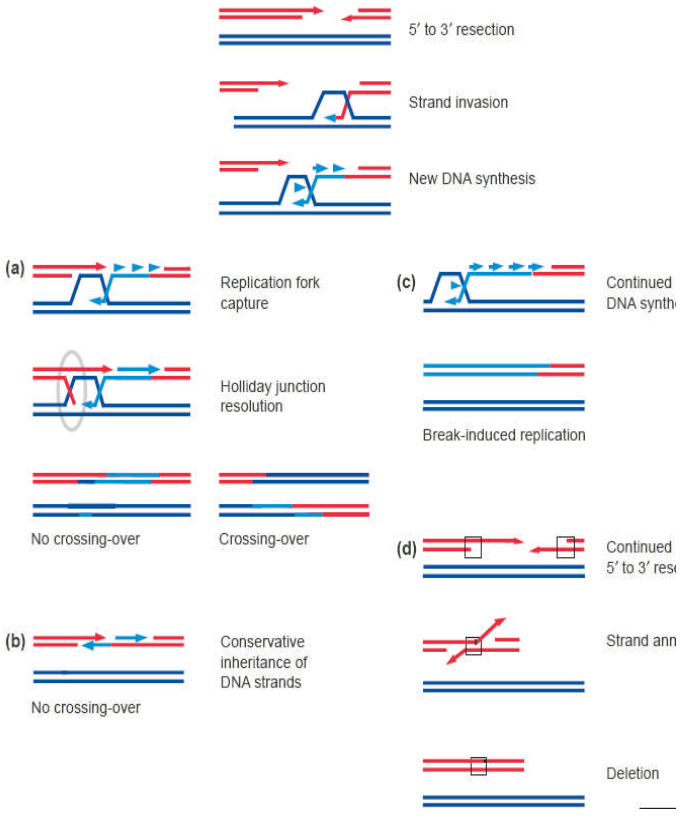


Figure 1. Proposed repair mechanisms for homologous recombination (Haber 2000).

At present, there are 4 models proposed for homologous recombinational repair: double strand break repair (DSBR), synthesis dependent strand annealing (SDSA), break induced repair (BIR) and single strand annealing (SSA) (Figure 1, a, b, c, and d, respectively). Studies on homologous recombination also provide the models for gene conversion.

After initiation by a chromosomal DSB, the processing of the DSB ends determines whether inaccurate repair or homologous recombination is used (HABER 2000). In yeast, if the overhangs at the

breakpoint are greater than 8 bases, the joining efficiency is greatly dependent on Rad52 (DALEY *et al.* 2005). Resection of the 5'end on the broken site leads to homologous recombination. Degradation of the 5' ends (resection) results in 3' overhangs on the two ends. Either one or two 3' protruding end(s) can invade the DNA template with the binding of RPA (Replication Protein A), a single-strand DNA binding protein, as well as Rad52 epistasis group proteins, a group of proteins that function sequentially and coordinate with each other during HR. The Rad52 epistasis group proteins (Rad51, Rad52, Rad54, Rad55, Rad57, Rad50, Mre11, and Xrs2 in yeast or Nbs1 in mammalian cells) are involved in mitotic recombination repair. There are very detailed reviews on the functions of these recombination proteins in yeast (KROGH and SYMINGTON 2004; SYMINGTON 2002). Additionally, in meiosis, Spo11, a topoisomerase II-like enzyme, is responsible for the formation of DSB (GRELON *et al.* 2001); and Dmc1, a meiosis-specific recombinase (BISHOP *et al.* 1992; MASSON and WEST 2001), is also involved in meiotic recombination.

In yeast, the resection of 5'ends is catalyzed by MRX (Mre11-Rad50-Xrs2) or by *Exo*I endonuclease, or other unidentified enzyme(s). In meiosis, MRX is suggested to be involved in the removal of *Spo11*, as well as the hairpin structure at the ends in DSB. RPA will coat the single strand DNA to remove the secondary structure. Rad52 interacts with RPA and facilitates the replacement of RPA by Rad51, which forms nucleoprotein filaments and will be associated with Rad55/57. Then the search for homologous sequences takes place. Once homology is found, the single strand will invade the donor DNA with the help of Rad54, a protein involved in homology search and strand invasion. After strand invasion, Rad51 is removed from DNA. DNA replication is primed by the invading 3' end of the broken strand. DSBR and SDSA models require a primed DNA synthesis process. The SDSA model is different from the DSBR

in that the first newly synthesized strands is unwound from the template and reunites with the broken molecule, allowing the two newly synthesized strands to pair with each other. Finally, the resolution of the Holliday Junction(s), depending on how it is resolved, will result in either crossover products or a non-crossover outcome. Mismatch repair (MMR) machinery is also involved (LAFLEURIEL *et al.* 2007; PAQUES and HABER 1999; SUGAWARA *et al.* 1997). MMR machinery inhibits meiotic recombination between divergent sequences in yeast. The MMR proteins participate in the recombination also. Mutations in MMR genes increase the proportion of gene conversion events (INBAR and KUPIEC 1999; INBAR and KUPIEC 2000). It is suggested that MMR genes are involved in anti-recombination activity and prevent exchange between highly diverged sequences. For example, in *Arabidopsis*, a loss of AtMSH2 increases the recombination frequency 3 fold (LAFLEURIEL *et al.* 2007).

No matter what mechanism is employed for the repair of breaks, an essential feature of homologous recombination is the formation of heteroduplex DNA followed by repair or correction, resulting in a non-reciprocal information transfer during gene conversion and also resulting sometimes in a crossover product. For gene conversion, the repair process almost always favors the donor strand (unbroken template) to be used as template in repair.

Besides DSBR and SDSA, SSA and BIR are the other two homologous recombination mechanisms. SSA will only happen when recombination occurs between two tandem repeat sequences, resulting in the deletion of the sequence between the repeats. The BIR model propose that after one 3' overhang invades the donor molecule, the primed DNA synthesis continues all the way to that invaded chromosome end, resulting in a long gene conversion tract.

Non-homologous end-joining

Ku70, Ku80, DNA-dependent protein kinase (DNA PK), Artemis (intrinsic single strand 5'to 3' exonuclease activity), DNA ligase IV, Xrcc4 and Xrcc-Like Factor (XLF) are involved in inaccurate repair in mammalian cells. In addition to that, in lower eukaryotes and in the V(D)J recombination in mammalian cells, the MRN(MRX in yeast) complex is also required for NHEJ (CLATWORTHY *et al.* 2005; MOORE and HABER 1996). MRX possesses the endonuclease activity that may be involved in the removal of the end structure and facilitates the Ku and ligase Dnl4 function in yeast (DALEY *et al.* 2005). In inaccurate repair, the Ku70/K80 heterodimer will first bind to both the DSB ends. It functions both as a bridge to the two ends, and attracts the DNA PK catalytic subunit to the break site with the help of Artemis. By phosphorylation, the DNA PK activates DNA ligase IV, and attracts other adapter proteins, such as Xrcc4, and XLF in mammalian cells. The DNA ligase IV joins the broken ends finally. The DNA end-binding protein, Ku70/80, binds to double-strand (ds) DNA during NHEJ. In homologous recombination, the end binding proteins such as RPA, as well as the Rad52 and Rad51 strand annealing and exchange proteins, bind to single-strand (ss) DNA end, and also bind the double-strand. Binding Ku heterodimers in NHEJ competes with the 5' resection and Rad52 binding in homologous recombination.

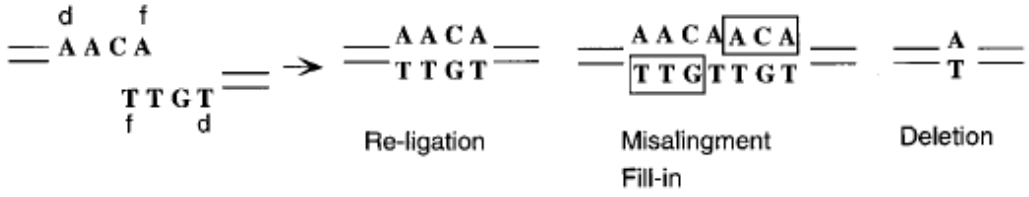


Figure 2. Inaccurate repair at HO endonuclease-cleaved DSB site. The resulting 4 bp overhangs can be relegated directly (Re-ligation). If the two 'f' marked bases pair, a 3 bp insertion will be created (Fill-in). If the two 'd' marked bases pair, a 3 bp deletion will be created (Deletion) (PAQUES and HABER 1999) .

As early as 2003, George Ilialis's lab identified an alternated NHEJ pathway in human cells, which is Ku-independent, DNA-PK independent (WANG *et al.* 2003). It was named as B (backup)-NHEJ, in order to distinguish it with D (DNA-PK dependent)-NHEJ. Now more features of this pathway have been uncovered. It is independent of Rad52 epistasis group proteins, and dependent on DNA ligase III, PARP-1(Poly ADP-ribose polymerase 1), and Histone H_1(TERZOUDI *et al.* 2008; WANG *et al.* 2006). In *Arabidopsis*, this pathway was also characterized and it was dependent on the proteins in chromatin structure maintenance (KOZAK *et al.* 2009). Also this pathway can complement the DSB repair when defects happened in the D-NHEJ. This research provides new insights into DSB repair mechanisms.

The differences between illegitimate recombination repair and homologous recombination in plants are summarized in Table 1.

Table 1. Comparisons between illegitimate recombination and homologous recombination in plants

Illegitimate Recombination	**Homologous Recombination**
Most active in somatic DNA repair	Most active in meiotic recombination
Preferred in G1 phase of cell cycle It can happen throughout the whole cell cycle.	Preferred for DNA repair in S and G2 phases of cell cycle
Episodic in response to DNA breaks	Pre programmed for normal segregation during meiosis, episodic for DNA repair
Error prone, allows nucleotide alterations at the sites of rejoining Sequence information lost due to deletions or insertions	High fidelity, eliminates aberrations Sequence information restored
No complementary DNA strand as template or micro-homology (< 9 nucleotides) between the breaking strand and the template strand (SDSA-like)	Complementary DNA strand as template
None or very little (<9 nucleotides) homology	Long stretch of homology (>30 nucleotides)
Direct ligation, or SDSA-like mechanism due to microhomology	Models proposed: DSBR, SDSA, SSA, and BIR
Ku70, Ku80, DNA-PK, Artemis, DNA ligase IV, XRCC *et al.* MRN complex (Mre11-Rad50-Nbs1)	RPA, Rad52 epistatis group proteins (Rad51, Rad52, Rad54, Rad55, Rad57, Mre11-Rad50-Nbs1) Spo11 and Dmc1 in meiosis only

The two pathways are complementary and competitive. The contributions of each pathway depends on the point in the cell cycle, the developmental stage and interspecies differences (PASTINK *et al.* 2001). In lower eukaryotes, such as yeast, homologous

recombination is the principal DSB repair mechanism. Illegitimate recombination has been found to be the prominent pathway in the somatic cells of the higher eukaryotes, mammals and plants (OREL and PUCHTA 2003; SANKARANARAYANAN and WASSOM 2005). In plants, a Rad52 homolog has not been found, which may partly explain why inaccurate repair, instead of homologous recombination, is the prominent pathway of DSB repair in plants (BOYKO *et al.* 2006b). Inaccurate repair can happen throughout all the stages of the cell cycle, especially during G0/G1, while homologous recombination is preferred during S and G2 phases when a complementary sister template is available. The preferences in DSB repair in plants are organ specific (BOYKO *et al.* 2006a), and are subject to developmental control (BOYKO *et al.* 2006b) and environmental factors, such as temperature and day length (BOYKO *et al.* 2005). The highest recombination frequency was found in the lateral half of leaves correlating with the level of metabolic activity (BOYKO *et al.* 2006a). As the plants grow older, the frequency of inaccurate repair is increased and the frequency of homologous recombination is decreased till the inaccurate repair becomes the principal DSB repair mechanism (BOYKO *et al.* 2006b). The decrease of homologous recombination corresponds to the increase of mutation frequency from 2 days post germination to 31 days post germination in *Arabidopsis* (BOYKO *et al.* 2006b). However, a higher mutation rate in tobacco corresponds to a higher recombination rate when compared with *Arabidopsis* in another report (FILKOWSKI *et al.* 2004).

2.2 DSB Repair Proteins and Plant Genome Stability

Because DSBs are common intrinsic intermediates in several cellular processes and they can also be easily induced by many extrinsic factors, cells must be able to repair DSBs correctly and efficiently. The direct role of DSB repair is to maintain genome integrity. A significant

feature of DSB repair is whether there is a complementary template strand available for the broken chromosome. Homologous recombination is relatively error free, while illegitimate recombination is error prone. Any deletions or insertions at the broken ends might lose or interrupt the original sequence information. Depending on where the DSB happens, changes in gene expression patterns, creation of novel alleles, or large chromosomal rearrangements are all possible. Deletion and insertion at a repair site will cause either a decrease or increase in the inherited genome size if this change happens in cells that contribute to the germ line. This may contribute significantly to the explanation of genome size variation in evolution.

DSBs can happen in both germ line cells and somatic cells. In germ line cells, defects in DSB repair will affect meiotic recombination and will often lead to serious reproduction defects. In somatic cells, genetic deficiencies in DSB repair will increase the X-ray sensitivity, but are usually viable with minor growing defects.

Research indicates that at least one recombination event per chromosome ensures proper segregation during the anaphase of meiosis. Mutations in any of the proteins in the repair pathways of DSB or the signaling pathways of chromosomal breaks checkpoint would result in different levels of genome instability. In *Arabidopsis*, proteins homologous to those in yeast that are involved in the inaccurate repair pathway have been found, such as AtKu70, AtKu80, AtLig4, AtXrcc4, AtMre11, and AtRad50, with protein functions varied through evolution. *Arabidopsis* also has homologous proteins in homologous recombination, such as AtDmc1, AtRad51B, AtRad51C, AtRad51D, AtXrcc2, and AtXrcc3, but it appears to lack a Rad52 homolog (OSAKABE *et al.* 2005). Rad52 is a key protein in homologous recombination. Furthermore, the AtKu70 levels, a protein involved in NHEJ pathway, but not AtRad51, strongly correlate with

the DSB levels in *Arabidopsis* (BOYKO *et al.* 2006b). This suggests that NHEJ is the primary DSB repair pathway in *Arabidopsis*.

Table 2. Plant mutants studied for their effects on DSB repair

Mutants	Plant phenotypes	References
Dmc1	Expressed in pollen mother cells and megaspore mother cells in ovules in *Arabidopsis*. Defects in bivalent formation. Subsequent unequal chromosome segregation and irregular spore generation, and induced changes in male meiotic progression in rice.	(COUTEAU *et al.* 1999; DENG and WANG 2007; KLIMYUK and JONES 1997)
Spo11	Required for initiation of meiosis and normal synapsis. Mutants in *Arabidopsis* have morphology similar to wild type but with significantly reduced seed numbers.	(GRELON *et al.* 2001)
Rad50	Higher expression in flowers and dividing cells. Mutants in *Arabidopsis* are sterile and hypersensitive to methyl methanesulfonate (MMS).	(BLEUYARD *et al.* 2004; GALLEGO *et al.* 2001; GALLEGO and WHITE 2001)
Rad51	Required for meiotic homology recognition, affecting meiotic chromosome pairing, synapsis, crossovers and normal meiotic disjunction, while dispensable for vegetative development. Mutants in *Arabidopsis* are completely sterile. In maize, mutant are male sterile and have significantly reduced seed number. Mutants have reduced numbers of bivalents and chiasmata, reduced homology pairing and increased non-homologous synapsis in meiosis. This gene is essential for the repair of radiation-induced DSBs during vegetative growth.	(LI *et al.* 2007; OSAKABE *et al.* 2002)
Rad51B	More highly expressed in buds and flowers, essential for mitosis. Mutant plants are smaller than the wild type, but viable and fertile with increased sensitivity to radiation.	(OSAKABE *et al.* 2005)
Rad51C	Required for meiotic synapsis. Mutants are completely male and female sterile. Normal vegetative and flower development without detectable abnormality in mitosis.	(LI *et al.* 2005)
Xrcc3	Required for meiosis but does not affect homologous chromosome synapsis in *Arabidopsis*. Mutants are sterile with normal vegetative development.	(BLEUYARD and WHITE 2004)
Rad54	More highly expressed in flower buds. Mutants are viable and sterile, increasing sensitivity to γ irradiation and cisplatin. Reduced homologous recombination efficiency in *Arabidopsis.*	(OSAKABE *et al.* 2006)
Mre11	Mutants are viable with growth defects, completely sterile with abundant dicentric chromosomes. Extensive fragmentation of chromosomes and absence of synapsis during meiosis in pollen mother cells. No effect on chromosomal fusions. Partially compensated by *spo11*$^{-}$ in *Arabidopsis*.	(PUIZINA *et al.* 2004)
Nbs1	*Arabidopsis* mutants exhibit normal growth and are fertile but show increased sensitivity to mitomycin C. Nbs1 plays a role in DNA damage signaling and the checkpoint pathway. The double mutant of AtNbs1-1 and AtATM (Ataxia-Telangiectasis Mutated) is sterile, with incomplete chromosome pairing and synapsis and chromosome fragmentation in metaphase of meiosis.	(WATERWORTH *et al.* 2007)
Ku70 and Ku80	NHEJ pathway, forming functional heterodimer with Ku80, binds to the DSB ends. Possessing ATP dependent helicase activity for both ssDNA and dsDNA in *Arabidopsis*. Important for activation of DNA-PK complex. Mutants have increased generation of DSB after exposure to bleomycin and MMS. No report on the fertility of Ku mutant plants yet..	(WEST *et al.* 2002)·(TAMURA *et al.* 2002)
Xrcc4	ArXrcc4 binds to the BRCT domain of *Arabidopsis* DNA ligase IV, and transcription increased under irradiation treatment.	(WEST *et al.* 2000)
DNA IV ligase	Enzyme involved in NHEJ. Induced by γ irradiation, interacts with Xrcc4.	(WEST *et al.* 2000)

RAD51 in DSB repair

Rad51 is a eukaryotic homologue of RecA, the bacterial recombinase that has a DNA-dependent ATPase activity. Yeast Rad51 was shown to form right-hand helical polymers on

both the dsDNA and ssDNA in an ATP-dependent manner and extend the DNA conformation (SUNG and ROBBERSON 1995). Rad51 belongs to the Rad52 epistasis group proteins, which include main factors in homologous recombination repair and meiotic recombination. Besides Rad51, the Rad51 family members (homologous to RecA) also include Dmc1 (meiotic specific recombinase), Rad55 and Rad57. Rad55 and Rad57 are involved in the stabilization of the Rad51-ssDNA complex in yeast.

Rad51 is involved in homologous chromosome pairing (PAWLOWSKI *et al.* 2003) in the meiotic recombination pathway. The early recombination nodules (RNs) are hypothesized to mark the non-crossover sites. RNs disappear between zygotene and late pachytene. It is believed that Rad51 is a component of early RNs (BISHOP 1994). The late RNs, which are densely stained and associated with tripartite synaptonemal complex (SC), exist throughout pachytene. From their distribution and frequency, it is believed that late RNs mark crossover sites (BISHOP 1994).

Rad51 has been shown to be involved in both somatic and meiotic DSB repair in *Drosophila* (YOO and MCKEE 2004; YOO and MCKEE 2005). It is dispensable for vegetative development but indispensable in meiosis in *Arabidopsis* (LI *et al.* 2004). Rad51 is involved in maize meiotic homologous recombination (PAWLOWSKI *et al.* 2003), and chromosome synapsis (FRANKLIN *et al.* 2003). *rad51* mutants in different species are varied in their phenotypes. For example, in *Drosophila*, *rad51* females are viable but sterile, while no effect is seen on male fertility and X-Y segregation in male meiosis. In *Arabidopsis*, *rad51* plants are viable but completely sterile due to their defects in both male and female meiotic recombination (LI *et al.* 2004). In maize, homozygous double mutants in two Rad51 homologues (*ZmRad51A1* and *ZmRad51A2*) are male sterile, and show reduced production of female gametes. But the

surviving female gametes were similar to the wild type in meiotic crossover frequency (LI *et al.* 2007). The species variation of *rad51* mutant phenotypes serves to underline the importance of studying Rad51 in different species in order to elucidate its function.

Besides its role in meiotic recombination, Rad51 functions in maintaining genome integrity. Rad51 displays a dynamic redistribution in nuclear foci at the DSB site in somatic cells. The nuclear distributions of Rad51 initiated by DSBs are closely related to the sites of chromosomal DSBs sites in maize (PAWLOWSKI *et al.* 2003), and in mammalian cells (MLADENOV *et al.* 2006). The level of Rad51 increases up to several folds in response to X-ray and methyl methanesulfonate (MMS) treatment. The heterozygous *rad51* mutants are more sensitive to X-ray and MMS treatment compared to wild type. In maize, the frequency of deletion-associated repair events increases up to 40 fold in the homozygous *rad51* double mutants (there are two copies of this gene) relative to the wild type in *MuDR*-induced DSBs (LI *et al.* 2008). The *rad51* mutants increase genomic instabilities in somatic cells. Studies on inaccurate repair events in maize *rad51* mutants help to explain the relationship between homologous recombination and inaccurate repair.

Rad51 paralogs

In mammals and *Arabidopsis*, five paralogs of Rad51 have been identified with 20-30% sequence identities with Rad51, and they are Rad51B, Rad51C, Rad51D, Xrcc2, and Xrcc3 (GODTHELP *et al.* 2002; OSAKABE *et al.* 2002). The expression levels of Rad51 family genes were higher in reproductive tissues, such as the young flower buds, than in vegetative tissues. Rad51, Rad51C and Xrcc3 were found to be highly expressed in roots but low in stems and leaves among the vegetative tissues in *Arabidopsis* (OSAKABE *et al.* 2002). The function of Rad51B in *Arabidopsis* is believed to be similar to its role in mouse, except that the *Arabidopsis*

rad51b mutant is viable without a growth defect phenotype, while the homozygous mouse *rad51b* mutant is lethal in an early embryonic stage (OSAKABE *et al.* 2005). In *Arabidopsis*, the Rad51C/Xrcc3 complex was shown to be required for meiosis, and the Rad51B/C/D/Xrcc2 complex may function in somatic cells (OSAKABE *et al.* 2005). The C-terminus of AtRad51B interacts with the C-terminus of AtRad51C in *Arabidopsis*. Xrcc3 interacts with Rad51 and Rad51Cα (OSAKABE *et al.* 2002).

Mre11 and Nbs1 in DSB repair

Mre11 and Nbs1, together with Rad50, form the MRN(X) (Mre11-Rad50-Nbs1) protein complex, which is involved in broken end resection in both inaccurate repair and homologous recombination, and is also involved in the removal of Spo11 in meiotic recombination. It has been shown that MRN affects synapsis of homologous chromosomes in *Arabidopsis* (PUIZINA *et al.* 2004) and the checkpoint in DSB repair. Mre11 and Rad50 both have ATPase activities that are important for tethering on DNA and a strong DNA bind activity is achieved with the help of Nbs1 (DALEY *et al.* 2005).

The role of the MRN complex in DSB repair is complicated. It has been reported that it is required for inaccurate repair in mammalian cells *in vitro* in addition to Ku, DNA protein kinase, and the DNA ligase/Xrcc4 complex (HUANG and DYNAN 2002). Mutations in *Mre11* and *Nbs1* are responsible for the Ataxia Telangiectasia-Like Disorder (ATLD) (STEWART *et al.* 1999) and the Nijmegen Breakage Syndrome (NBS), respectively (VARON *et al.* 1998). The association of MRN complex activities with the ATLD and NBS alleles provided evidence that MRN plays a role in inaccurate repair. However, hypomorphic mutations in *mre11* or *nbs1* did not show clear DSB repair deficiencies in human or mouse cells (DI VIRGILIO and GAUTIER 2005).

In addition to this, although a 3' to 5' nuclease activity is associated with Mre11 *in vivo*, the Mre11 protein was shown to have a 5' to 3' activity only *in vitro*. A structural role of the complex that bridges the two DNA ends during DSB repair was more pronounced than its nuclease activity in yeast mitotic DSB repair (DUDASOVA *et al.* 2004). The MRX complex in yeast may assist Ku and Dnl4 during non-homologous end-joining (NHEJ) (DALEY *et al.* 2005).

Due to its complicated and fundamental roles, multiple phenotypes are seen in *mre11* mutants. It is difficult to separate its function in somatic DSB repair from its role in meiotic recombination in higher eukaryotes. The sterility of *mre11* homozygous mutants poses another difficulty in studying its function in plant. The *Arabidopsis mre11* homozygous T-DNA insertion mutants are viable but with multiple growth defects, and are infertile due to extensive fragmentations of chromosomes (PUIZINA *et al.* 2004). Abundant dicentric chromosomes and chromosomal fragments can also be observed during mitosis. Study of the *ku70/80* and *mre11* mutants did not show a reduced frequency of chromosomal fusion, implying that the plants have a Ku/Mre11-independent DNA end-joining activity. Studies on the *spo11* and *mre11* double mutants showed a partially recovered sterility compared with the *mre11* mutant and suppressed chromosomal fragmentation, implying that Mre11 participates in the Spo11-initiated DSB repair during meiosis in *Arabidopsis* (PUIZINA *et al.* 2004).

Nbs1 is one of the three components of the MRN complex. Nbs1 is redistributed to the DSB sites together with Mre11/Rad50 upon exposure to ionizing radiation in human cells (SOMEYA *et al.* 2006). *Arabidopsis* and maize Nbs1 are homologues to mammalian Nbs1 (WATERWORTH *et al.* 2007). Nbs1 has the FHA and BRCT binding domain at the N-terminal end, and the Mre11 binding domain at the C-terminus. Deletion of the C-terminal domain results in the accumulation of Mre11/Rad50 in the cytoplasm instead of entering into the nucleus in

response to irradiations. The functions of Nbs1 in DSB repair and meiotic recombination have been studied in yeast, mammalian cells, and plant cells, but the exact roles are not clear. Nbs1 has two ATM (Ataxia-Telangiectasia Mutated) phosphorylation sites at Ser278 and Ser343 that are phosphorylated in the presence of Mre11/Rad50 with irradiation-induced DSBs (CEROSALETTI and CONCANNON 2004). This indicates a cell cycle checkpoint role of Nbs1. Nbs1 stimulates the DNA binding and nuclease activity of Mre11/Rad50. The roles of Nbs1 in plant cells need to be further elucidated.

2.3 DSB Repair Mechanisms And Chromosomal Rearrangement

Besides its main role in maintaining genome stability, DSB repair functions in shaping genome structure and size. So far, numerous predicted outcomes of inaccurate NHEJ events have been identified in plants, such as the *Ac/Ds* transposon excisions (GORBUNOVA and LEVY 2000; YAN *et al.* 1999), and within-transposon deletions (YAN *et al.* 1999), T-DNA integrations (GHEYSEN *et al.* 1991; MAYERHOFER *et al.* 1991; TZFIRA *et al.* 2003) and within-gene deletions, such as at the maize *Wx* allele (WESSLER *et al.* 1990). The huge variations created by inaccurate repair contribute significantly to gene evolution.

As mentioned before, DSB repair pathways are highly conserved in different species, as are the effects of different DSB repair mechanisms on chromosome rearrangement. For inaccurate repair, either deletions or insertions are associated with the break repair site. The inserted sequence may either be apparently random or showing homology to a chromosome segment. Microhomologies (less than 4 nucleotides) between the cutting ends and nearby sequences are common for NHEJ (PAQUES and HABER 1999). The frequent small deletions

and/or insertions are the hallmark of illegitimate recombination (MA and BENNETZEN 2004). Large deletions or insertions are also possible but not as frequent as the small ones.

Among the four different potential pathways for homologous recombination, the DSBR model and the SDSA model types of recombination can generate both crossover and non-crossover events and result in gene conversions, leading to a non-Mendelian segregation ratio. A change in the allelic frequencies may have profound effects on genome evolution. SSA happens when recombination occurs between two repeat sequences and results in deletion of the sequence between the repeats and one copy of the repeat. Depending on the locations of two repeat segments, either at homologous or ectopic locations, SSA would result in different effects. The naturally-occurring events have been identified in *Arabidopsis* (DEVOS *et al.* 2002). In the BIR model, after one 3' overhang invades the donor molecule, DNA synthesis is primed and continues to the end of the invaded chromosome, thus copying the entire template strand. This has been observed between plasmid DNA and yeast chromosomes during an artificial transformation experiment.

Recombination is a prominent feature of both meiosis and somatic DSB repair. Chromosomal recombination is a very active research field because of its basic biological importance, such as in gamete generation, and its significant role in genome evolution. Nowadays, people have found more and more evidence of DSB-initiated recombination and its resultant changes on gene functions and structures. The idea that DSB repair contributes to genome size variation is widely accepted.

2.4 Comparative Study Of The Effect Of DSB Repair On Genome Evolution

Species-specific DSB repair events are related to genome evolution

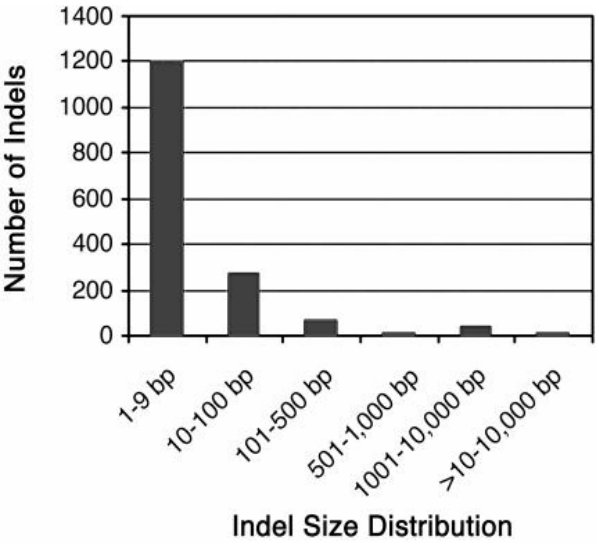


Figure 3. Small deletions, possibly caused by DSB repair, were identified in haplotypic comparison across the rice genome (MA *et al.* 2004).

How inaccurate repairs affect genome structure and size is a fascinating topic. Large-size deletions appear to be rare in plant DSB repair events, perhaps as a protective function to avoid the loss of important genes. The nature of DSB repairs in maintaining genome integrity is closely related with cells' survival ability under natural selection and environmental adaptation. Also, this small deletion pattern of inaccurate repairs has been identified in other species, such as in human cells (HONMA *et al.* 2007).

DSB repair plays a significant role in genome evolution (BENNETZEN 1998; BENNETZEN 2002; BENNETZEN 2007; BENNETZEN *et al.* 2005; KIRIK *et al.* 2000; MA *et al.* 2004). Frequent small deletions (1-9 nucleotides) that have been suspected to be caused by DSB repair exist throughout the rice genome in both *indica* and *japonica* subspecies, when compared with the African rice, *Oryza glaberrima* (MA *et al.* 2004).

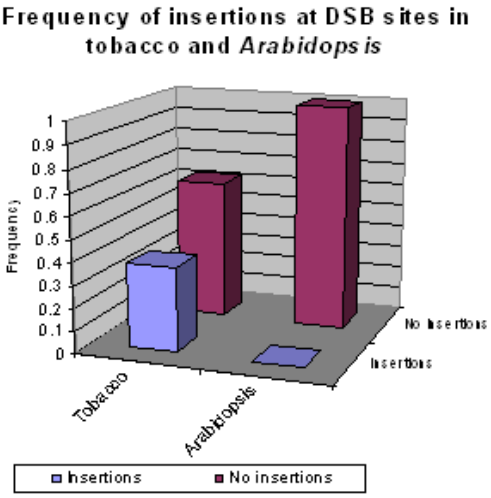


Figure 4. The frequency of insertions and deletions at double-strand breaks identified in *Arabidopsis* and tobacco (KIRIK *et al.* 2000).

The Hohn and Puchta labs were the first to study the outcomes of engineered double strand break repair in plants (PUCHTA *et al.* 1993; PUCHTA *et al.* 1996). Comparative studies on the I-*Sce* I-induced site-specific DSB repair events in *Arabidopsis* and tobacco, which has a 20 times larger genome size than *Arabidopsis,* indicate that all the DSB repair events in *Arabidopsis* are associated with deletions, with no insertions identified. In tobacco, up to 40% of the DSB repair events were associated with insertions. The differences of DSB repair were postulated to help explain the 20-fold variation in the genome sizes between *Arabidopsis* and tobacco, two eudicot species (KIRIK *et al.* 2000; SALOMON and PUCHTA 1998).

In another study on *Arabidopsis* and tobacco, the authors found that the homologous recombination rate in tobacco is 75 fold higher than that in *Arabidopsis,* and proposed that tobacco utilize homologous recombination to repair DSBs more frequently than *Arabidopsis* (FILKOWSKI *et al.* 2004). So far, there are no data available concerning the relative rates of inaccurate repair frequency in these two species, so more detailed conclusions are not possible.

How species perform DSB repair may be closely related with genome size change, but this issue needs more experimental evidence. In my thesis research, I used the same I-*Sce* I enzyme as Dr. Puchta's lab used to introduce DSBs in maize, rice, sorghum and pearl millet, grasses with very different genome sizes. For example, rice is ~400 Mbp, maize ~2500 Mbp, and sorghum ~780 Mbp. The genome size of peal millet was estimated to be ~2450 Mbp (2n = 14), with a C-value of 2.36 pg (ALLOUIS *et al.* 2001; MARTEL *et al.* 1997). By comparing inaccurate repair events among these monocots, and between the monocots and the dicots, we may be able to draw conclusions on the effect of DSB repair on genome size evolution.

Extrinsic factors affect genome evolution

Recombination is one of the most dominant features involved in genome evolution (GAUT *et al.* 2007). Studies in plants could uncover more details about how DSB repairs, especially homologous recombination, affect genome evolution. It was indicated that recombination frequency is negatively correlated with metabolic rate and positively correlated with the concentration of peroxide produced in plants (BOYKO *et al.* 2005). However, this conclusion was not in agreement with the observations in another study done by the same lab, in which a higher somatic recombination rate in lateral halves of the leaves corresponded to a higher metabolic rate compared with the medial halves of leaves (BOYKO *et al.* 2006a). The recombination rate measured on the recombined reporter transgene activity (β-glucuronidase or luciferase) was observed to be negatively correlated with the day length in one study, possibly because of higher transgene activity and higher endoreduplication levels in plants grown at longer days, with a 15-fold difference in the recombination frequency observed between plants grown at the shortest and the longest days (BOYKO *et al.* 2005). The incongruent observations under different conditions complicate the study on the mechanisms of DSB repairs in plants. Another interesting study suggested that the low-frequency of loss of intact BARE-1 retrotransposons by unequal homologous recombination was correlated with a stressful environment (high altitude and dryness) (KALENDAR *et al.* 2000), but this study did not control for population history.

In addition to the effect on genome size, DSB repairs may play an important role in the evolution of gene function. In response to the stressful environment, such as disease or day length changes, genomic DNA may evolve by adaptation and natural selection. Studies on cereal *Vrn1* loci in wheat and barley indicates that spring *Vrn1* alleles are derived from winter alleles

illegitimate recombination (COCKRAM *et al.* 2007). In the analysis of the different mutant alleles of maize waxy gene (*Wx*), it was found that spontaneous insertions ranging from 1 to 131 bp happened between the deletions endpoints, and those deletion termini were clustered within a 1000 bp region, a hot spot characterized as having a high GC content and significantly lower free energies to stabilize secondary structure formations in physiological conditions (WESSLER *et al.* 1990). It was not known whether the inaccurate DNA repair plays a significant role in *Wx* evolution, but the nature of the haplotype variability is suggestive of such a role.

2.5 Research Methods Inducing DSBs In Plants

In this research, I utilized I-*Sce* I, a rare-cutting endonuclease encoded by the group I intron of the mitochondrial LSU gene of *Saccharomyces cerevisiae*. This "homing" endonuclease was first discovered during analysis of the ω genetic system of yeast mitochondria. In this research, it was used to introduce DSB in maize, rice and other cereal genomes. It has a long (18 bp) recognition site (Figure 5), which has not yet been found to be represented in any plant genome (data not shown). Assuming random organization of the sequence, the frequency of occurrence of 18 bp sites is one per 4^{18} (6.87×10^{10}), which is equal to one site per 70, 000 Mbp, about 25 times the size of the maize genome, and 20 mammalian-sized genomes (JASIN 1996). Cleavage by I-*Sce* I produces 3' overhangs of four bases. In this research, we obtained the I-*Sce* I ORF sequence from Dr. Puchta.

5'-ATTACCCTGTTAT↓CCCTA-3'
3'-TAATGGGAC↑AATAGGGAT-5'

Figure 5. The recognition and cleavage sites of the I-*Sce* I endonuclease.

Besides I-*Sce* I, other endonucleases such as the HO rare-cutting endonucleases , zinc-finger proteins (LLOYD *et al.* 2005), and transposase (LI *et al.* 2008) have been used to induce DSBs in plants, as well as other organisms.

In this research, two T-DNA constructs were made: one has the I-*Sce* I recognition site flanked by the cauliflower 35s promoter (*p35S*) and *CodA* gene; the other has the I-*Sce* I coding region under the control of the maize *Ubi*-1 promoter and *nos* 3' terminator (CHRISTENSEN and QUAIL 1996). The two T-DNA segments were inserted into the pTF101.1 binary vector, respectively. The target site T-DNA construct and the I-*Sce* I ORF were obtained from Dr. Holger Puchta, who had been using I-*Sce* I to introduce DSBs via transient transformation in *Arabidopsis* and tobacco cells (SALOMON and PUCHTA 1998). The *Ubi*-1 promoter sequence and the *nos* 3' terminator were obtained from Dr. Peter Quail (CHRISTENSEN and QUAIL 1996). The Plant Transformation Facility of Iowa State University generated these transformant lines with the constructs that I assembled, putting them into maize and rice. The sorghum and pearl millet transgenic lines were recently made by Dr. Thomas Clemente at the Core Research Facility of the University of Nebraska in Lincoln.

2.6 Significance

DSB repair is a very important chromosomal behavior. This project is the first to study I-*Sce* I-induced DNA repair events in a monocot species. This will also be the first study to determine whether there are differences in DNA repair within plant species at different genomic locations, or whether repair is performed differently at different times in somatic development. The study of DSB repairs in maize mutants is meaningful in filling the many gaps in our current knowledge of these DNA repair genes in plants.

However, this study is not limited to these points. The hallmark of this study is that we are using an experimental method to test and confirm our hypothesis that inaccurate repairs play an important role in shaping genomes by characterizing and comparing the inaccurate repair events in four different monocots that have varied genome sizes. Since the first time inaccurate DNA replication or repair was proposed to be a genome shrinkage mechanism in *Drosophila* by Petrov and colleagues (PETROV *et al.* 1996), this phenomena have been studied in dicot species, *Arabidopsis* and tobacco (FILKOWSKI *et al.* 2004; KIRIK *et al.* 2000). Experimental evidence from monocot species is provided in this study. This allows us to draw more accurate conclusions regarding the effects of DSB repair on genome evolution by comparing with the data from dicots.

This study focuses on grass family species, which are closely related but with huge genome size variations. Using the monocot species as our study model gives some benefits. First, rice is used as a model monocot plant. It has the one of the smallest genome sizes and its full genome sequence is available to the public. Second, the extensive synteny among the grasses (DEVOS 2005; DEVOS and GALE 1997) like rice, maize, barley, wheat, sorghum, and several millets, provides additional benefits of comparative genomic study. Third, since many agriculturally important plants belong to this family, the resultant high priority of their sequencing projects makes either whole genome sequences and/or BAC libraries available. This is helpful in finding the chromosomal locations of target site T-DNA insertions, as needed in the study of any possible position effects on DSB repair.

2.7 References

ALLOUIS, S., X. QI, S. LINDUP, M. D. GALE and K. M. DEVOS, 2001 Construction of a BAC library of pearl millet, *Pennisetum glaucum*. Theor and Appl Genet **102:** 1200-1205.

BENNETZEN, J. L., 1998 The structure and evolution of angiosperm nuclear genomes. Curr Opin Plant Biol **1:** 103-108.

BENNETZEN, J. L., 2002 Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica **115:** 29-36.

BENNETZEN, J. L., 2007 Patterns in grass genome evolution. Curr Opin in Plant Biol **10:** 176-181.

BENNETZEN, J. L., J. MA and K. M. DEVOS, 2005 Mechanisms of recent genome size variation in flowering plants. Ann Bot **95:** 127-132.

BISHOP, D. K., 1994 RecA homologs Dmc1 and Rad51 interact to form multiple nuclear complexes prior to meiotic chromosome synapsis. Cell **79:** 1081-1092.

BISHOP, D. K., D. PARK, L. XU and N. KLECKNER, 1992 DMC1: A meiosis-specific yeast homolog of *E. coli* recA required for recombination, synaptonemal complex formation, and cell cycle progression. Cell **69:** 439-456.

BLEUYARD, J. Y., M. E. GALLEGO and C. I. WHITE, 2004 Meiotic defects in the *Arabidopsis* rad50 mutant point to conservation of the MRX complex function in early stages of meiotic recombination. Chromosoma **113:** 197-203.

BLEUYARD, J. Y., and C. I. WHITE, 2004 The *Arabidopsis* homologue of Xrcc3 plays an essential role in meiosis. Embo J **23:** 439-449.

BOYKO, A., J. FILKOWSKI, D. HUDSON and I. KOVALCHUK, 2006a Homologous recombination in plants is organ specific. Mutat Res **595:** 145-155.

BOYKO, A., J. FILKOWSKI and I. KOVALCHUK, 2005 Homologous recombination in plants is temperature and day-length dependent. Mutat Res **572:** 73-83.

BOYKO, A., F. ZEMP, J. FILKOWSKI and I. KOVALCHUK, 2006b Double-strand break repair in plants is developmentally regulated. Plant Physiol **141:** 488-497.

BREE, R. T., C. NEARY, A. SAMALI and N. F. LOWNDES, 2004 The switch from survival responses to apoptosis after chromosomal breaks. DNA Repair (Amst) **3:** 989-995.

CEROSALETTI, K., and P. CONCANNON, 2004 Independent roles for nibrin and Mre11-Rad50 in the activation and function of Atm. J Biol Chem **279:** 38813-38819.

CHRISTENSEN, A. H., and P. H. QUAIL, 1996 Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res **5:** 213-218.

CLATWORTHY, A. E., M. A. VALENCIA-BURTON, J. E. HABER and M. A. OETTINGER, 2005 The MRE11-RAD50-XRS2 complex, in addition to other non-homologous end-joining factors, is required for V(D)J joining in yeast. J Biol Chem **280:** 20247-20252.

COCKRAM, J., I. J. MACKAY and D. M. O'SULLIVAN, 2007 The role of double-stranded break repair in the creation of phenotypic diversity at cereal VRN1 loci. Genetics **177:** 2535-2539.

COUTEAU, F., F. BELZILE, C. HORLOW, O. GRANDJEAN, D. VEZON *et al.*, 1999 Random Chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc1* mutant of *Arabidopsis*. Plant Cell **11:** 1623-1634.

DALEY, J. M., P. L. PALMBOS, D. WU and T. E. WILSON, 2005 Nonhomologous end joining in yeast. Annu Rev Genet **39:** 431-451.

DENG, Z. Y., and T. WANG, 2007 OsDMC1 is required for homologous pairing in *Oryza sativa*. Plant Mol Biol **65:** 31-42.

DEVOS, K. M., 2005 Updating the 'crop circle'. Curr Opin Plant Biol **8:** 155-162.

DEVOS, K. M., J. K. BROWN and J. L. BENNETZEN, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res **12:** 1075-1079.

DEVOS, K. M., and M. D. GALE, 1997 Comparative genetics in the grasses. Plant Mol Biol **35:** 3-15.

DI VIRGILIO, M., and J. GAUTIER, 2005 Repair of double-strand breaks by nonhomologous end joining in the absence of Mre11. J Cell Biol **171:** 765-771.

DUDASOVA, Z., A. DUDAS and M. CHOVANEC, 2004 Non-homologous end-joining factors of *Saccharomyces cerevisiae*. FEMS Microbiol Rev **28:** 581-601.

FILKOWSKI, J., O. KOVALCHUK and I. KOVALCHUK, 2004 Dissimilar mutation and recombination rates in *Arabidopsis* and tobacco. Plant Science **166:** 265-272

FRANKLIN, A. E., I. N. GOLUBOVSKAYA, H. W. BASS and W. Z. CANDE, 2003 Improper chromosome synapsis is associated with elongated RAD51 structures in the maize desynaptic2 mutant. Chromosoma **112:** 17-25.

GALLEGO, M. E., M. JEANNEAU, F. GRANIER, D. BOUCHEZ, N. BECHTOLD *et al.*, 2001 Disruption of the *Arabidopsis* RAD50 gene leads to plant sterility and MMS sensitivity. Plant J **25:** 31-41.

GALLEGO, M. E., and C. I. WHITE, 2001 RAD50 function is essential for telomere maintenance in *Arabidopsis*. Proc Natl Acad Sci U S A **98:** 1711-1716.

GAUT, B. S., S. I. WRIGHT, C. RIZZON, J. DVORAK and L. K. ANDERSON, 2007 Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet **8:** 77-84.

GHEYSEN, G., R. VILLARROEL and M. VAN MONTAGU, 1991 Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev. **5:** 287-297.

GODTHELP, B. C., F. ARTWERT, H. JOENJE and M. Z. ZDZIENICKA, 2002 Impaired DNA damage-induced nuclear Rad51 foci formation uniquely characterizes *Fanconi* anemia group D1. Oncogene **21:** 5002-5005.

GORBUNOVA, V., and A. A. LEVY, 2000 Analysis of extrachromosomal *Ac/Ds* transposable elements. Genetics **155:** 349-359.

GRELON, M., D. VEZON, G. GENDROT and G. PELLETIER, 2001 AtSPO11-1 is necessary for efficient meiotic recombination in plants. Embo J **20:** 589-600.

HABER, J. E., 2000 Partners and pathways: repairing a double-strand break. Trends Genet **16:** 259-264.

HONMA, M., M. SAKURABA, T. KOIZUMI, Y. TAKASHIMA, H. SAKAMOTO *et al.*, 2007 Non-homologous end-joining for repairing I-*Sce*I-induced DNA double strand breaks in human cells. DNA Repair **6:** 781-788.

HUANG, J., and W. S. DYNAN, 2002 Reconstitution of the mammalian DNA double-strand break end-joining reaction reveals a requirement for an Mre11/Rad50/NBS1-containing fraction. Nucleic Acids Res **30:** 667-674.

INBAR, O., and M. KUPIEC, 1999 Homology search and choice of homologous partner during mitotic recombination. Mol Cell Biol **19:** 4134-4142.

INBAR, O., and M. KUPIEC, 2000 Recombination between divergent sequences leads to cell death in a mismatch-repair-independent manner. Curr Genet **38:** 23-32.

JASIN, M., 1996 Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet **12:** 224-228.

KALENDAR, R., J. TANSKANEN, S. IMMONEN, E. NEVO and A. H. SCHULMAN, 2000 Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci U S A **97:** 6603-6607.

KIRIK, A., S. SALOMON and H. PUCHTA, 2000 Species-specific double-strand break repair and genome evolution in plants. Embo J **19:** 5562-5566.

KLIMYUK, V. I., and J. D. JONES, 1997 AtDMC1, the *Arabidopsis* homologue of the yeast DMC1 gene: characterization, transposon-induced allelic variation and meiosis-associated expression. Plant J **11:** 1-14.

KOZAK, J., C. E. WEST, C. WHITE, J. A. DA COSTA-NUNES and K. J. ANGELIS, 2009 Rapid repair of DNA double strand breaks in *Arabidopsis thaliana* is dependent on proteins involved in chromosome structure maintenance. DNA Repair **8:** 413-419.

KROGH, B. O., and L. S. SYMINGTON, 2004 Recombination proteins in yeast. Annu Rev Genet **38:** 233-271.

LAFLEURIEL, J., F. DEGROOTE, A. DEPEIGES and G. PICARD, 2007 Impact of the loss of AtMSH2 on double-strand break-induced recombination between highly diverged homeologous sequences in *Arabidopsis thaliana* germinal tissues. Plant Mol Biol **63:** 833-846.

LI, J., L. C. HARPER, I. GOLUBOVSKAYA, C. R. WANG, D. WEBER *et al.*, 2007 Functional analysis of maize RAD51 in meiosis and double-strand break repair. Genetics **176:** 1469-1482.

LI, J., T.-J. WEN and P. S. SCHNABLE, 2008 Role of RAD51 in the repair of *MuDR*-induced double-strand breaks in maize (*Zea mays L.*). Genetics **178:** 57-66.

LI, W., C. CHEN, U. MARKMANN-MULISCH, L. TIMOFEJEVA, E. SCHMELZER *et al.*, 2004 The *Arabidopsis* AtRAD51 gene is dispensable for vegetative development but required for meiosis. Proc Natl Acad Sci U S A **101:** 10596-10601.

LI, W., X. YANG, Z. LIN, L. TIMOFEJEVA, R. XIAO *et al.*, 2005 The AtRAD51C gene is required for normal meiotic chromosome synapsis and double-stranded break repair in *Arabidopsis*. Plant Physiol. **138:** 965-976.

LLOYD, A., C. L. PLAISIER, D. CARROLL and G. N. DREWS, 2005 Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. Proc Natl Acad Sci U S A **102:** 2232-2237.

MA, J., and J. L. BENNETZEN, 2004 Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci U S A 101: 12404-12410.

MA, J., K. M. DEVOS and J. L. BENNETZEN, 2004 Analyses of *LTR*-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res **14:** 860-869.

MARTEL, E., D. DE NAY, S. SILJAK-YAKOVIEV, S. BROWN and A. SARR, 1997 Genome size variation and basic chromosome number in pearl millet and fourteen related *Pennisetum* species. J Hered **88:** 139-143.

MASSON, J.-Y., and S. C. WEST, 2001 The Rad51 and Dmc1 recombinases: a non-identical twin relationship. Trends Biochem Sci **26:** 131-136.

MAYERHOFER, R., Z. KONCZ-KALMAN, C. NAWRATH, G. BAKKEREN, A. CRAMERI *et al.*, 1991 T-DNA integration: a mode of illegitimate recombination in plants. EMBO J **10:** 697-704.

MLADENOV, E., B. ANACHKOVA and I. TSANEVA, 2006 Sub-nuclear localization of Rad51 in response to DNA damage. Genes Cells **11:** 513-524.

MOORE, J. K., and J. E. HABER, 1996 Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature **383:** 644-646.

OREL, N., and H. PUCHTA, 2003 Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. Plant Mol Biol **51:** 523-531.

OSAKABE, K., K. ABE, H. YAMANOUCHI, T. TAKYUU, T. YOSHIOKA *et al.*, 2005 *Arabidopsis* Rad51B is important for double-strand DNA breaks repair in somatic cells. Plant Mol Biol **57:** 819-833.

OSAKABE, K., K. ABE, T. YOSHIOKA, Y. OSAKABE, S. TODORIKI *et al.*, 2006 Isolation and characterization of the RAD54 gene from *Arabidopsis thalian*a. Plant J **48:** 827-842.

OSAKABE, K., T. YOSHIOKA, H. ICHIKAWA and S. TOKI, 2002 Molecular cloning and characterization of RAD51-like genes from *Arabidopsis thaliana*. Plant Mol Biol **50:** 71-81.

PAQUES, F., and J. E. HABER, 1999 Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. **63:** 349-404.

PASTINK, A., J. C. EEKEN and P. H. LOHMAN, 2001 Genomic integrity and the repair of double-strand DNA breaks. Mutat Res **480-481:** 37-50.

PAWLOWSKI, W. P., I. N. GOLUBOVSKAYA and W. Z. CANDE, 2003 Altered nuclear distribution of recombination protein RAD51 in maize mutants suggests the involvement of RAD51 in meiotic homology recognition. Plant Cell **15:** 1807-1816.

PETROV, D. A., E. R. LOZOVSKAYA and D. L. HARTL, 1996 High intrinsic rate of DNA loss in *Drosophila*. Nature **384:** 346-349.

PUCHTA, H., B. DUJON and B. HOHN, 1993 Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. Nucleic Acids Res **21:** 5034-5040.

PUCHTA, H., B. DUJON and B. HOHN, 1996 Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci U S A **93:** 5055-5060.

PUIZINA, J., J. SIROKY, P. MOKROS, D. SCHWEIZER and K. RIHA, 2004 Mre11 deficiency in *Arabidopsis* is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. Plant Cell **16:** 1968-1978.

ROOS, W. P., T. NIKOLOVA, S. QUIROS, S. C. NAUMANN, O. KIEDRON *et al.*, 2009 Brca2/Xrcc2 dependent HR, but not NHEJ, is required for protection against O6-methylguanine triggered apoptosis, DSBs and chromosomal aberrations by a process leading to SCEs. DNA Repair **8:** 72-86.

SALOMON, S., and H. PUCHTA, 1998 Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. Embo J **17:** 6086-6095.

SANKARANARAYANAN, K., and J. S. WASSOM, 2005 Ionizing radiation and genetic risks: XIV. Potential research directions in the post-genome era based on knowledge of repair of radiation-induced DNA double-strand breaks in mammalian somatic cells and the origin of deletions associated with human genomic disorders. Mutat Res **578:** 333-370.

SOMEYA, M., K. SAKATA, H. TAUCHI, Y. MATSUMOTO, A. NAKAMURA *et al.*, 2006 Association of ionizing radiation-induced foci of NBS1 with chromosomal instability and breast cancer susceptibility. Radiat Res **166:** 575-582.

STEWART, G. S., R. S. MASER, T. STANKOVIC, D. A. BRESSAN, M. I. KAPLAN *et al.*, 1999 The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. Cell **99:** 577-587.

SUGAWARA, N., F. PAQUES, M. COLAIACOVO and J. E. HABER, 1997 Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand break-induced recombination. Proc Natl Acad Sci U S A **94:** 9214-9219.

SUNG, P., and D. L. ROBBERSON, 1995 DNA strand exchange mediated by a RAD51-ssDNA nucleoprotein filament with polarity opposite to that of RecA. Cell **82:** 453-461.

SYMINGTON, L. S., 2002 Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. Microbiol Mol Biol Rev **66:** 630-670

TAMURA, K., Y. ADACHI, K. CHIBA, K. OGUCHI and H. TAKAHASHI, 2002 Identification of Ku70 and Ku80 homologues in *Arabidopsis thaliana*: evidence for a role in the repair of DNA double-strand breaks. Plant J **29:** 771-781.

TERZOUDI, G. I., S. K. SINGH, G. E. PANTELIAS and G. ILIAKIS, 2008 Premature chromosome condensation reveals DNA-PK independent pathways of chromosome break repair. Int J Oncol **33:** 871-879.

TZFIRA, T., L. R. FRANKMAN, M. VAIDYA and V. CITOVSKY, 2003 Site-specific integration of *Agrobacterium tumefaciens* T-DNA via double-stranded intermediates. Plant Physiol **133:** 1011-1023.

VARON, R., C. VISSINGA, M. PLATZER, K. M. CEROSALETTI, K. H. CHRZANOWSKA *et al.*, 1998 *Nibrin*, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. Cell **93:** 467-476.

VITTE, C., and J. L. BENNETZEN, 2006 Eukaryotic Transposable Elements and Genome Evolution Special Feature: Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci U S A **103:** 17638-17643.

WANG, H., A. R. PERRAULT, Y. TAKEDA, W. QIN, H. WANG *et al.*, 2003 Biochemical evidence for Ku-independent backup pathways of NHEJ. Nucleic Acids Res **31:** 5377-5388.

WANG, M., W. WU, W. WU, B. ROSIDI, L. ZHANG *et al.*, 2006 PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res **34:** 6170-6182.

WATERWORTH, W. M., C. ALTUN, S. J. ARMSTRONG, N. ROBERTS, P. J. DEAN *et al.*, 2007 NBS1 is involved in DNA repair and plays a synergistic role with ATM in mediating meiotic homologous recombination in plants. Plant J **52:** 41-52.

WESSLER, S., A. TARPLEY, M. PURUGGANAN, M. SPELL and R. OKAGAKI, 1990 Filler DNA is associated with spontaneous deletions in maize. Proc Natl Acad Sci U S A **87:** 8731-8735.

WEST, C. E., W. M. WATERWORTH, Q. JIANG and C. M. BRAY, 2000 *Arabidopsis* DNA ligase IV is induced by gamma-irradiation and interacts with an *Arabidopsis* homologue of the double strand break repair protein XRCC4. Plant J **24:** 67-78.

WEST, C. E., W. M. WATERWORTH, G. W. STORY, P. A. SUNDERLAND, Q. JIANG *et al.*, 2002 Disruption of the *Arabidopsis* AtKu80 gene demonstrates an essential role for AtKu80 protein in efficient repair of DNA double-strand breaks in vivo. Plant J **31:** 517-528.

WICKER, T., N. YAHIAOUI and B. KELLER, 2007 Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. Plant J **51:** 631-641.

YAN, X., I. M. MARTINEZ-FEREZ, S. KAVCHOK and H. K. DOONER, 1999 Origination of *Ds* elements from *Ac* elements in maize: evidence for rare repair synthesis at the site of *Ac* excision. Genetics **152:** 1733-1740.

YOO, S., and B. D. MCKEE, 2004 Overexpression of *Drosophila* Rad51 protein (DmRad51) disrupts cell cycle progression and leads to apoptosis. Chromosoma **113:** 92-101.

YOO, S., and B. D. MCKEE, 2005 Functional analysis of the *Drosophila* Rad51 gene (spn-A) in repair of DNA damage and meiotic chromosome segregation. DNA Repair (Amst) **4:** 231-242.

CHAPTER 3

THE INTRODUCTION OF AN ENGINEERED DSB SYSTEM INTO SEVERAL GRASS SPECIES

In this chapter, I describe the methodology that I used to generate DSBs in the genomes of four grasses with very different genome sizes. The technology that was developed can be applied to any other grass species in future studies.

3.1 Construction Of T-DNA Binary Vectors

T-DNA binary vector constructs for the I-*Sce* I ORF and its 18 bp recognition site

The target site vector, pTFBNE, was generated by inserting the I-*Sce* I recognition site flanked by the *p35S* promoter and the *CodA* gene with *nos* 3' terminator into the appropriate cloning site of pTF101.1 (Figure 6). The pTF101.1 binary vector, provided by Iowa State University, was the carrier of the T-DNA sequence integrated into plant genomes through *Agrobacterium*-mediated plant transformation. The binary vector

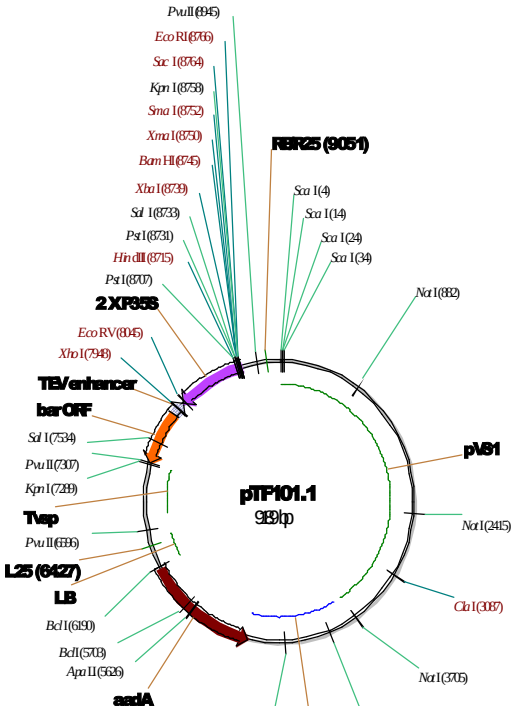
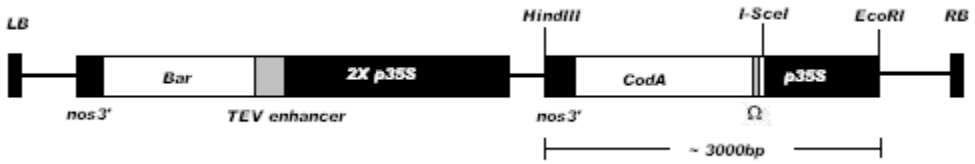


Figure 6. The pTF101.1 vector used for plant transformation.

pTF101.1 was chosen because it can be efficiently used in *Agrobacterium* bacteria-mediated transformation in monocots (FRAME *et al.* 2002) at Iowa State University. The I-*Sce* I recognition site was excised from the pBNE3I vector as an *Eco*RI and *Hin*dIII fragment of about 3 kb. Vector pBNE31 was obtained from Dr. Puchta (PUCHTA *et al.* 1996; SALOMON and PUCHTA 1998). The *p35S* promoter and *nos* 3' terminator driving *CodA* gene expression were present so that Puchta and coworkers could select for mutations that affected *CodA* expression. This was not a useful feature for my planned project, because the *CodA* selection is not robust in grasses (SALOMON and PUCHTA 1998).

The I-*Sce* I enzyme expression vector, pTFSceI, was generated in 3 steps. The pAHC17 vector was obtained from Dr. Quail. First, the *nos* 3' terminator sequence from the pAHC17 vector was cut by *Bam*HI and *Eco*RI (295 bp), and ligated into the pTF101.1 vector, thus generating pTFNOS. Second, the *Ubi*-1 promoter sequence was excised as a 2,017 bp *Bam*HI/*Hin*dIII fragment from plasmid pAHC17 (CHRISTENSEN and QUAIL 1996). This promoter from a maize ubiquitin gene was chosen to drive the I-*Sce* I ORF because it was found to be expressed at high levels in all tissues tested (CHRISTENSEN and QUAIL 1996) and in both maize and rice (CORNEJO *et al.* 1993). The *Ubi*-1 promoter fragment was ligated into pTFNOS, generating pTFNOSUBI. Finally, the I-*Sce* I ORF segment, which was cut from pCISCEI (SALOMON and PUCHTA 1998) with *Bam*HI, was ligated into pTFNOSUBI, generating pTFSceI. There were two possible insertion directions of the I-*Sce* I ORF, so I restriction mapped clones in order to select one with the I-*Sce* I ORF in the appropriate orientation so that it would be expressed under the *Ubi*-I promoter.

T-DNA with the enzyme recognition site (pTFBNE)



T-DNA with the insertion of I-*Sce*I ORF (pTFSceI)

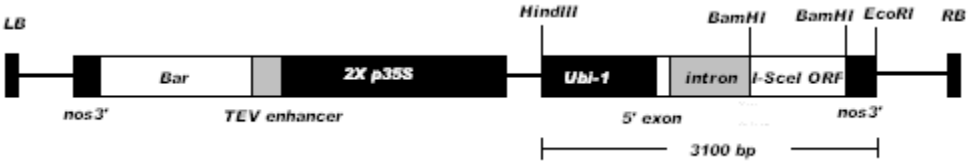


Figure 7. Schematic diagrams of the T-DNA components for the two binary vectors that were produced. *Bar* was the first selectable marker: seeds from parental transgenic plants were selected on medium containing phosphinothrycin (PPT, active ingredient of the herbicide Basta). *CodA*, the second selectable marker, is a gene encoding cytosine deaminase, but was not used for selection in this study. *p35s*, cauliflower mosaic virus 35s promoter. 2x*p35S*, the double strength cauliflower mosaic virus 35S promoter. *nos 3'*, the nopaline synthase terminator. The *Ubi-1* promoter fragment contains 899 bp of promoter sequence, 83 bp of 5' untranslated exon and 1010 bp of first intron sequence from the maize Ubiquitin-1 gene. *TEV*, tobacco etch virus translational enhancer. *Ω* (SALOMON and PUCHTA 1998). RB, T-DNA right border; LB, T-DNA left border.

Confirmation of the T-DNA binary vector structure by restriction enzyme digestion and sequencing

The structures of the two T-DNA vectors were confirmed by restriction enzyme digestion and by insert boundary sequencing. The sequence analysis also confirmed that there were no mutations at the engineered sites within the constructs. This is especially important for the vector pTFSceI with the I-*Sce* I ORF for the enzyme expression because there were two possible directions for the I-*Sce* I ORF insertions.

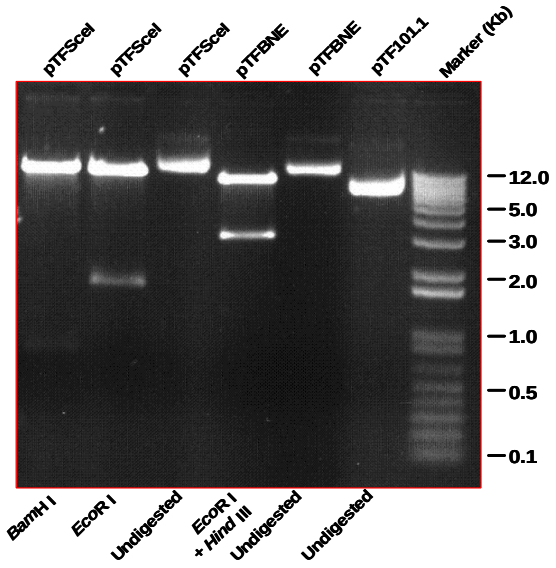


Figure 8. Confirmation of two T-DNA constructs by restriction enzyme digestion.

The insertions of these T-DNA constructs into the binary vectors were also confirmed through sequencing, using a Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). The primers used are listed in Table 3. Sequencing reactions were prepared as in Table 4. The sequencing program was 94 °C for 3 min; 39 cycles at 94°C for 30 sec, 58°C for 20 sec, and 60°C for 4 min; followed by 60°C for 8 min.

Table 3. Primers used for vector sequencing

Primers for sequencing pTFBNE		Primers for sequencing pTFScel	
pTF101R	GGCTCGTATGTTGTGTGGAAT	pTF101R	GGCTCGTATGTTGTGTGGAAT
S_{10} [a]	GCGGGAAGCTTCAGCTGACGCGTACACAACA	ISceI3'	TTGGAATTTGTTACGCAGACC
S_{11} [a]	GCGGGAAGCTTGAACTCGCCGTGAAGACTGG	tfsceR1	CGAACACTATCTCCTCCGAAA
BNEtar5	CGGGTCGACGGTACCTATTA	Ubiant1	AGATGACCCGACAAACAAGTG
BNEtar3	TCGACGGATCCTAGGGATAA	Ubiant2	CAGACCACATCATCACAACCA
A_4 [a]	CGCTCTAGAGCCCCTCTTCGCCTGGTAAC	Ubiant3	CAATTTCTGGATGCCGACAG
A_{10} [a]	GCGTCTAGACATTTTCAGCCGGCAGGATA	Ubiant4	AGGCTGGCATTATCTACTCGAA
A_1 [a]	TTGACTCTAGAGGATCAACG	pTFhin5	TGACCTGCAGGCATGCAAGCTT
pTFhin5	TGACCTGCAGGCATGCAAGCTT		

[a] The primer sequences were obtained from SALOMON, S., and H. PUCHTA, 1998. The other primers were designed by PRIMER3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).
* The grey shaded primers were used for sequencing of vector pTFBNE in the reverse direction.

Table 4. Big Dye sequencing reaction conditions

	1 X Reaction
H_2O	3.5 µl
5X Sequencing Buffer	1 µl
Primer (5 pmol / µl)	1.5 µl
Big Dye	1 µl
DNA template	3 µl

All of the sequencing performed within this thesis research was analyzed on the ABI3730 that is available at the UGA Center of Research Equipment. The vector sequences are in Appendix A.

3.2 Generation Of Transgenic Lines Of Maize, Rice, Sorghum, And Pearl Millet

These transformation experiments resulted in two types of transgenic plants: one with the I-*Sce* I recognition site and another with I-*Sce* I enzyme expression. For maize and rice, these transformants were all generated by the Iowa State University Plant Transformation Facility (http://www.agron.iastate.edu/ptf/). For maize, constructs were transformed into line Hi II, yielding lines called A112 with the I-*Sce* I target site (81 plantlets from 12 transgenic events) and lines called A113 with the I-*Sce* I ORF (79 plantlets from 12 transgenic events). For rice, constructs were transformed into rice cultivar *Nipponbare*, generating lines called R32 with the I-*Sce* I target site (16 transgenic events, 95 plantlets) and lines called R33 with the I-*Sce* I ORF (12 transgenic events, 127 plantlets). Transgenic seedlings obtained from Iowa State University were transplanted to soil and allowed to fully mature.

The T-DNA constructs for sorghum and peal millet were the same but were carried on different binary vectors: pPTN819 with the target site and pPTN820 with the I-*Sce* I expression construct. These constructs were assembled by Dr. Thomas Clemente at the University of Nebraska in Lincoln, who also performed all of the transformation. The transgenic sorghums and pearl millets were obtained as T0 seeds instead of plantlets. For sorghum, lines with the

target site were generated from 9 different events while lines with I-*Sce* I ORF were generated from 4 different events. Thirty sorghum parent plants from 9 different events with the target site insertion and 17 plants from 4 different events with the I-*Sce* I ORF insertion were planted and screened for the insertion using the methods described in chapter 3.3. The transgenic pearl millets were also available but they were not included in my thesis research due to time limitations.

3.3 Genotyping Of Parental Transgenic Lines In Maize, Rice, And Sorghum

All of the parental transgenic plants were screened to confirm the presence of the appropriate T-DNA insertions before crosses were made for generating the hemizygous recombinant progenies that would induce DSBs through I-*Sce* I digestion.

Materials and methods

Sample collection

About 1 month after planting the maize and rice plantlets into soil in the greenhouse, 20 to 25 leaf punches from one or two leaves were sampled using a standard-size paper puncher (0.7 mm in diameter) for each plant. The leaf disks were collected directly into 1.5 ml centrifuge tubes. The samples were frozen in liquid nitrogen and stored at -80°C before freeze-drying.

DNA isolation

Two steel beads were added to the tubes to grind the leaf tissue into a powder by vibrant shaking by a Mix Mill MM330 (Qiagen). Total genomic DNA from the leaf samples was isolated according to the protocol on http://www.agron.iastate.edu/ptf/service/analysis.aspx with 2X CTAB as extraction buffer and without phenol extraction. The genomic DNA was dissolved in 50 µl of TE buffer (10 mM Tris-HCl, 1mM EDTA, pH = 8.0).

PCR

PCR was done using primers specific to each of the T-DNA insertions (S_{11} and A_4 for the target site, and ISceI5'- CCGTTCTCGTGATGAAGGTAA and ISceI3'-TTGGAATTTGTTACGCAGACC for the enzyme ORF). The touchdown PCR program used was 94°C for 1 min, 6 cycles at 94°C for 30sec, 64°C for 25 sec, 64-57°C for 50 sec (decrease 1°C per cycle); 34 cycles of 94°C for 20 sec, 57°C for 25 sec, 72°C for 50 sec; and 72°C for 10 min, then 4°C. The results confirmed that about 95% of the maize and rice transformants received from Iowa State University were carrying T-DNA insertions (Figure 9).

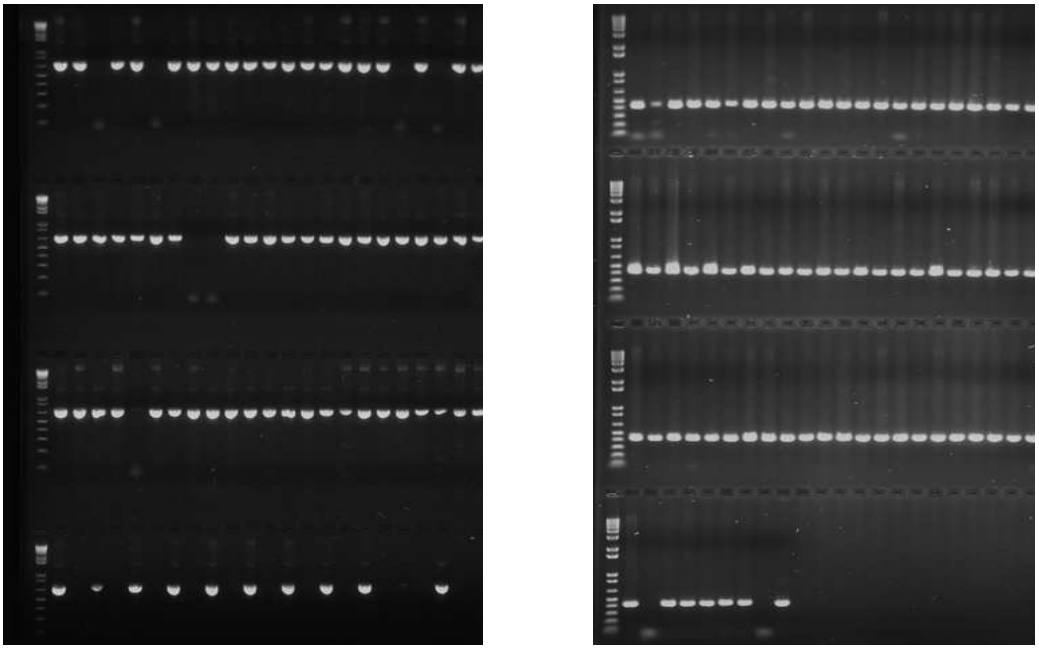


Figure 9. PCR confirmation of the pTFBNE T-DNA insertions in A112 lines in maize with the target site, and pTFSceI T-DNA insertions in the A113 lines in maize with the I-*Sce* I ORF. About 95% of maize and rice transformants were confirmed to be carrying T-DNA insertions.

RT-PCR

For the transgenic lines with the I-*Sce* I ORF driven by *Ubi*-I promoter, RT-PCR was done with the same pair of primers to confirm the expression of I-*Sce* I enzyme. RNA was extracted from leaf tissue using Trizol (Invitrogen). RT-PCR was done using the Invitrogen

SuperScript III First-Strand Synthesis SuperMix followed by regular PCR using primers specific to the I-*Sce* I ORF: ISceI5' and ISceI3'. As can be seen in Figure 10, most of the I-*Sce* I ORF transgenes were expressed.

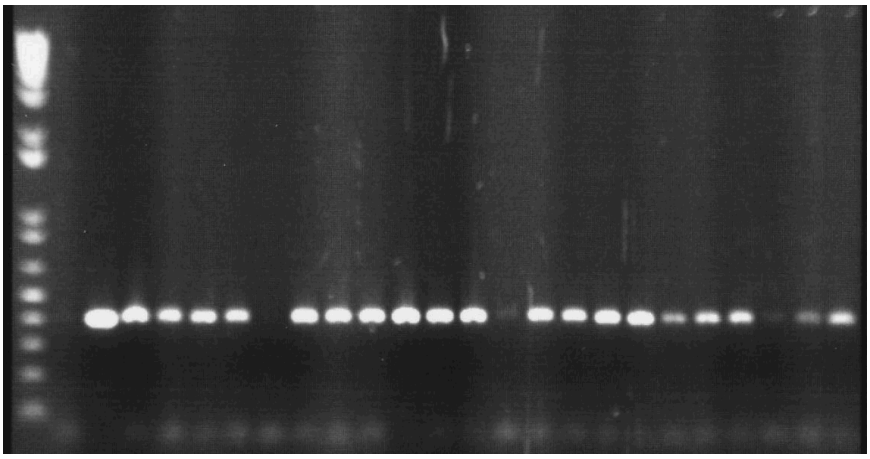


Figure 10. RT PCR confirmation of the expression of the I-*Sce* I ORF in the maize A113 lines.

3.4 Induction Of DSBs By Crosses Between The Two Transgenic Lines

Individual plants from target site insertion lines were crossed with individual plants from enzyme expression lines to obtain F_1 seed. A variety of crosses were made, depending on the availability of the pollen and female flowers during development.

For maize, both the T-DNA insertion lines were also maintained via crosses to the B73 inbred. A total of 864 maize F_1, 161 rice F_1, and 99 sorghum F_1 were produced. In these F_1 individuals, the expression of I-*Sce* I should initiate a site-specific cut where the enzyme recognition site was integrated into the genome.

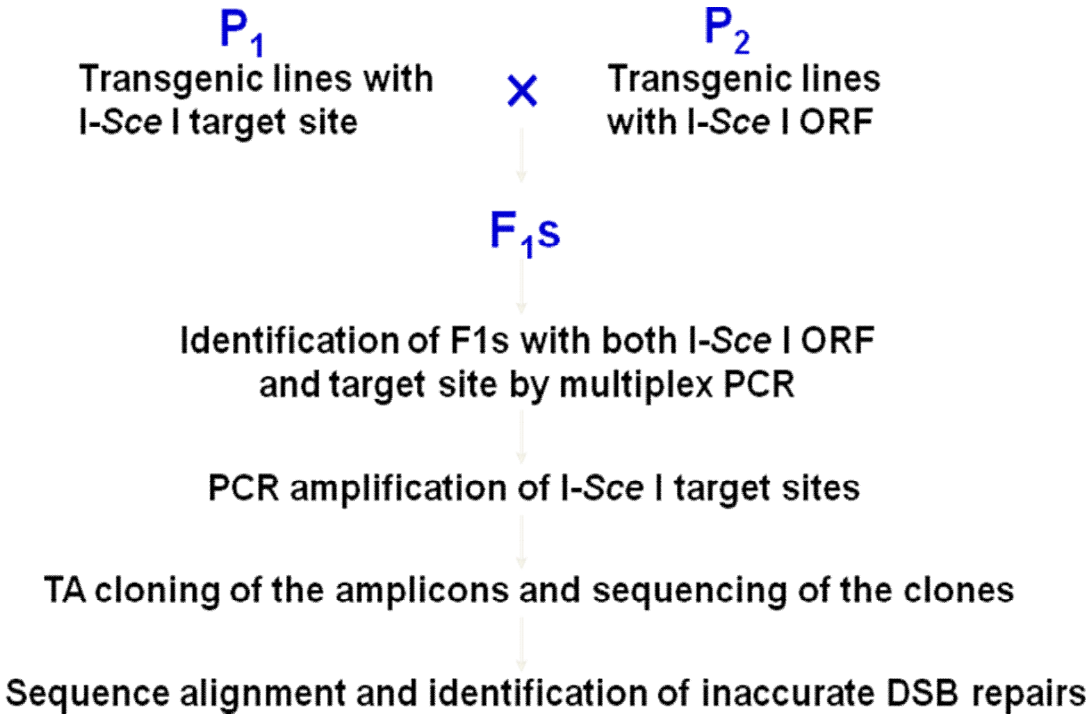


Figure 11. Diagram of the experimental design.

CHAPTER 4

MAPPING THE LOCATIONS OF T-DNA INSERTIONS IN MAIZE AND RICE

4. 1 Isolation Of Genomic Sequences Flanking T-DNA Insertions By TAIL-PCR

TAIL(thermal asymmetric interlaced)-PCR uses two or three nested-specific primers in consecutive reactions together with an arbitrary degenerate (AD) primer having a lower Tm (melting temperature), so that the relative amplification efficiencies of specific and non-specific products can be thermally controlled (LIU and HUANG 1998). TAIL-PCR was used to amplify the flanking sequences of inserted target sites. Specific primers facing outward from the T-DNA ends were designed and paired with an AD primer to amplify both the left T-DNA border (LB) and right T-DNA border (RB) insertion site, followed by a second round of PCR with one nested specific primer and the AD primer. Ten to twenty nanograms of genomic DNA was used as template. Genomic sequences flanking the T-DNA insertions were amplified according to Liu *et al.* (LIU and WHITTIER 1995).

Table 5. The primers used in TAIL-PCR

Arbitrary degenerate primers		Specific primer to T-DNA borders	
RiceAD8 [a]	(G/C)TTGNTA(G/C)TNCTNTGC	tailRB1	ACACAACATACGAGCCGGAAG
RiceAD9 [a]	(A/T)CAGNTG(A/T)TNGTNCTG	tailRB2	CATACGAGCCGGAAGCATAAA
MaizAD0 [b]	NTCGA(G/C)T(A/T)T(G/C)G(A/T)GTT	tailRB3	GGTTTGCGTATTGGAGCTTGA
MaizAD1 [b]	NGTCGA(G/C)(A/T)GANA(A/T)GAA	tailLB1	TACCCAACTTAATCGCCTTGC
MaizAD6 [b]	(A/T)GTGNAG(A/T)ANCANAGA	tailLB2	AGCTGGCGTAATAGCGAAGAG
MaizeW4 [c]	AG(A/T)GNAG(A/T)ANCANAGA	tailLB3	CCTGAATGGCGAATGCTAGAG
		tailLB4	GCAATTCGGCGTTAATTCAGT

[a] Sha et al 2004 (SHA *et al.* 2004) ; [b] Liu et al 1995 (LIU *et al.* 1995) ; [c] Settles et al 2004 (SETTLES *et al.* 2004).

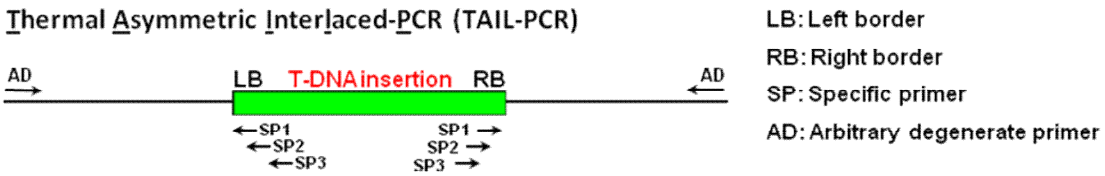


Figure 12. Diagram of the TAIL-PCR method used in this project.

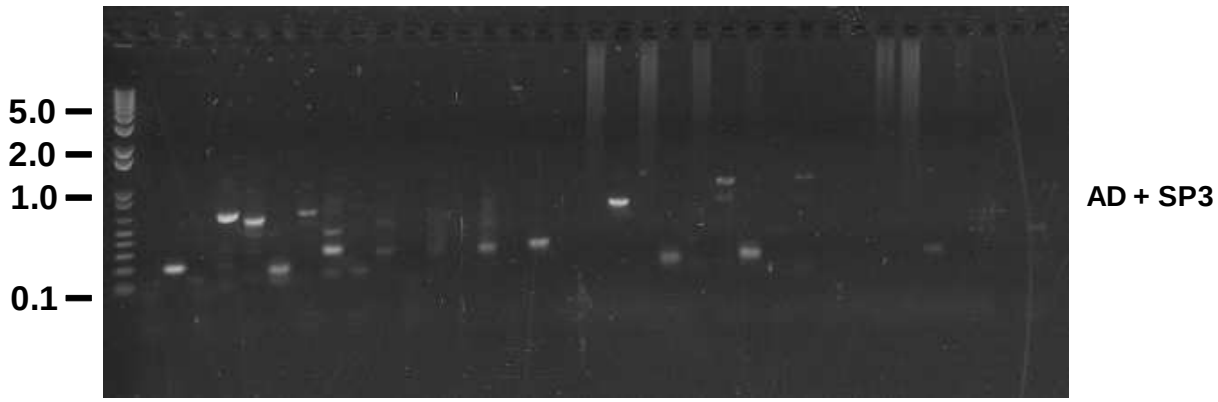


Figure 13. An example of TAIL-PCR products from lines A112 and lines R32 after the third round PCR. Different lanes represent the different amplifications using different AD primers paired with specific primers either for left border or right border. Samples are from independent events of transgenic lines for maize (A112) and rice (R32).

Table 6. TAIL-PCR thermocycling conditions

Primary reaction		
Step	Temperature	Time
1	94	2 min
2	95	1min
3	94	20 seconds
4	62	1 min
5	72	2 min 30 seconds
6	Go to Step 3	5 times
7	94	20 seconds
8	25	3 min
9	Ramp to 72	+ 0.3 ℃/seconds
10	72	2 min 30 seconds
11	94	20 seconds
12	62	1 min
13	72	2 min 30 seconds
14	94	20 seconds
15	62	1 min
16	72	2 min 30 seconds
17	94	20 seconds
18	44	1 min
19	72	2 min 30 seconds
20	Go to Step 11	14 times
21	72	5 min
22	4	End

Secondary reaction		
Step	Temperature	Time
1	94	2 min
2	94	20 seconds
3	62	1 min
4	72	2 min 30 seconds
5	94	20 seconds
6	62	1 min
7	72	2 min 30 seconds
8	94	20 seconds
9	44	1 min
10	72	2 min 30 seconds
11	Go to step 2	11 times
12	72	5 min
13	4	End

Tertiary reaction		
Step	Temperature	Time
1	94	2 min
2	94	20 seconds
3	44	1 min
4	72	2 min 30 seconds
5	Go to step 2	30
6	4	End

To determine the positions of target site T-DNA insertions, the amplified 200-700 bp of T-DNA flanking sequences were cloned and sequenced. The sequences (Appendix C) were used to search NCBI (http://www.ncbi.nlm.nih.gov/mapview/static/MVPlantBlast.shtml?4577) for nucleotide similarities by the BLASTN algorithm. All the rice T-DNA border sequences obtained were able to be specifically mapped to a unique chromosome. However, BLASTN using the maize T-DNA border sequences to the NCBI databases usually resulted in no significant similarity. The http://www.maizesequence.org, http://www.plantsequnce.org and http://www.tigr.org/plantProjects.shtml were used to find the neighboring genomic background for maize target site T-DNA insertions.

To determine whether any of these insertions interrupted genes or were found inside repeats, cDNA databases and repeat databases were searched using the flanking sequences by the BLASTN algorithm.

4.2 Target Sites Mapped In The Maize Genome

Table 7. *In-silico* mapping of the target sites in the maize genome by BLAST

Event	Genome Location	Annotation
A112-15	AC206877.1_Contig42 Chromosome 4	*Zea mays* putative *Fourf* gag/pol protein
A112-16	AC194174.3-Contig45 Chromosome 7	*Zea mays Fourf* gag/pol protein
A112-17	AC200869.2:1-177895 Chromosome 5	Similarity to transposable elements
A112-20 *	AC210165.1-Contig27 Chromosome 8, 4, 10	Similarity to 22-kDa alpha zein gene cluster
A112-4 *	AC200659.3 Chromosome 6	Hypothetical protein in *Oryza*
A112-1 *	AC206313 Chromosome 6	Putative gag/pol precursor
A112-14 *	AC148110.2 Chromosome 5	Putative reverse trascriptase in *Zingiber officinalis* (Ginger)
A112-18 *	AC194174.3-Contig45 Chromosome 4	*Zea mays* putative *Fourf* gag/pol protein

* indicates the hits with highest score, but the chromosomal locations were not unique or the annotation was in process.

For rice, 12 of the 16 independent target site transformation events were characterized and the features flanking the target sites for all 12 were hypothetical proteins. The For maize, since 70-80% of genome is non-gene coding sequence, I expected higher probabilities that the independent target site insertions were located in repeated sequences. At present, 8 independent events have been characterized in maize: 4 were classified as putative gag/pol sequence, 2 were transposable element related, 1 putative reverse transcriptase, and 1 was annotated as a candidate gene for a hypothetical protein. For 5 of the independent maize target sites, BLAST failed to give a specific location, probably due to the currently incomplete genome database. The T-DNA flanking sequences need further check in detail.

4.3 Target Sites Mapped In The Rice Genome

The T-DNA border sequences obtained from TAIL-PCR were searched by BLAST against the NCBI database (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=4530) of *Oryza Sativa*, to map the T-DNA target site insertions. The output is summarized in Table 8.

Twelve target site T-DNA insertion in rice were successfully mapped on chromosome 1, 3, 5 ,7, 8, 9, 11. Among all the target sites being mapped, 4 target sites were mapped on the chromosome 8, and 3 were mapped on the chromosome 3. T-DNA insertions toward Chromosome 8 and 3 appeared more frequent than expected.

Table 8. *In-silico* mapping of the target sites in the rice genome by BLAST

Event	Genome Location	Annotation
R32-7	Chromosome 5	4,233 bp at 5' side: hypothetical protein 1,492 bp at 3' side: hypothetical protein
R32-8	Chromosome 1	24,377 bp at 5' side: hypothetical protein 484 bp at 3' side: hypothetical protein
R32-9	Chromosome 9	249 bp at 5' side: hypothetical protein 4,515 bp at 3' side: hypothetical protein
R32-12	Chromosome 8	984 bp at 5' side: hypothetical proteins 9,173 bp at 3' side: hypothetical proteins
R32-13	Chromosome 3	267 bp at 5' side: hypothetical proteins 5,584 bp at 3' side: rRNA-45S ribosomal RNA
R32-15	Chromosome 8	14,829 bp at 5' side: hypothetical protein 16,910 bp at 3' side: hypothetical protein
R32-18	Chromosome 3	8,708 bp at 5' side: hypothetical proteins 10,851 bp at 3' side: hypothetical proteins
R32-20	Chromosome 8	4,572 bp at 5' side: hypothetical proteins 3,358 bp at 3' side: hypothetical proteins
R32-21	Chromosome 3	1,382 bp at 5' side: hypothetical proteins 1,2249 bp at 3'side: hypothetical proteins
R32-22	Chromosome 8	4,506 bp at 5' side: hypothetical proteins 3,355 bp at 3' side: hypothetical protein
R32-25	Chromosome 11	12,025 bp at 5' side: hypothetical protein 24,223 bp at 3' side: hypothetical protein
R32-32	Chromosome 7	10,513 bp at 5' side: hypothetical proteins 17,661 bp at 3' side: hypothetical proteins

CHAPTER 5

CHARACTERIZATION OF INACCURATE DSB REPAIR IN MAIZE AND RICE

5.1 Screening F_1 To Find Individuals With Both I-*Sce* I ORF And Its Target Site

Because both the I-*Sce* I ORF and I-*Sce* I target site transgenics were expected to be hemizygous for the presence of the transgene, a cross between two of these lines will only yield 25% of the progeny with both of the necessary components to promote the desired DSB events. Hence, twelve or more F_1 seed from each cross were planted if enough F_1 seed were collected.

Multiplex-PCR was used to identify those hemizygous F_1 progenies with both an I-*Sce* I ORF insertion and its target site. In such lines, DSBs should occur due to I-*Sce* I expression and the presence of its target cleavage site in these F_1 progenies. If accurate repair occurs, either by homologous recombination or NHEJ, the restored enzyme recognition site will be cut again until an inaccurate repair occurs. The break site cut by I-*Sce* I that was not repaired at all would block the cell cycle and cause the cell to die. Therefore, eventually, all repair events identified would be inaccurate repair events.

A total of 220 maize F_1, 47 rice F_1, and 24 sorghum F_1 were identified as having both the enzyme and its target site by multiplex PCR. A segregation ratio of 1 (both the target site and the enzyme ORF) to 2 (either the target site or the enzyme but not both) to 1 (no insertions at all — wild type) was expected and observed, indicating that most of the *Agrobacterium* mediated T-DNA insertions were single copy in each T_0 transgenic plant.

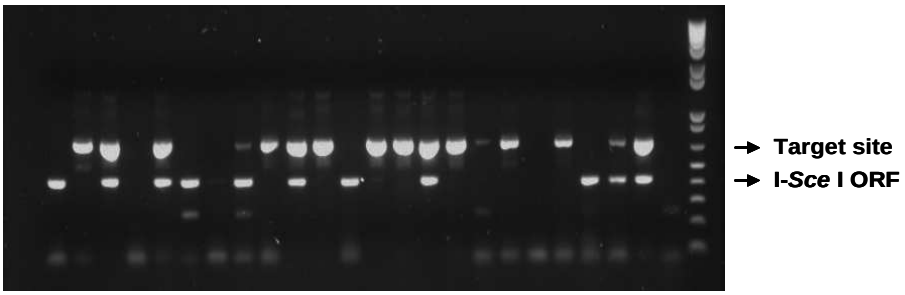


Figure 14. Multiplex PCR was used to identify F_1 with both insertions. Two primer pairs were used. One primer pair (S_{11} and A_4) was used to amplify the target site; the other primer pair (ISce5' and ISce3') was used to amplify the I-*Sce* I ORF.

5.2 Target Site-Specific PCR Amplification, Cloning And Sequencing

Primers flanking the DSB site were used to amplify the DNA sequence at the target site using the DNA extracted from leaf. The original target site sequence was used as control, and it gave an amplification product of about 650 bp with the primer pairs that were employed. The PCR amplification products were cloned into the TOPO TA vector (Invitrogen TOPO cloning kit) for sequencing. For each amplicon, at least 8 clones were randomly selected for sequencing. A comprehensive description of the inaccurate repair events will be acquired in this way. A total of 192 clones were sequenced for maize, 384 clones were sequenced for rice, and 192 clones were sequenced for sorghum.

Because PCR was performed on DNA from leaf tissues, and the timing of inaccurate repair was not initially known, there was no way to predict at first whether clones from a specific F_1 would have all the same inaccurate repair event or would have different events because of somatic sectoring. By characterization the repair events identified in different clones from one individual, I could acquire some understanding of how often different somatic repair events happened in one individual plant. When the same sequence was found in different clones from the same individual, this was interpreted as the outcome of a single event, although there was

also a significant possibility that some of these could have been derived from separate occurrences of the identical event. Hence, quantization of the relative frequencies of events was not perfect, and would under-represent the most common class of inaccurate DSB repair outcomes. Figure 15 presents an example of the length polymorphism of the PCR products due to inaccurate repairs identified by the PCR analysis of individual colonies that were derived from target site PCR amplicons of a single F_1 plant. This confirmed the success in our methods of using I-*Sce* I to induce an inaccurate site-specific DSB repair event.

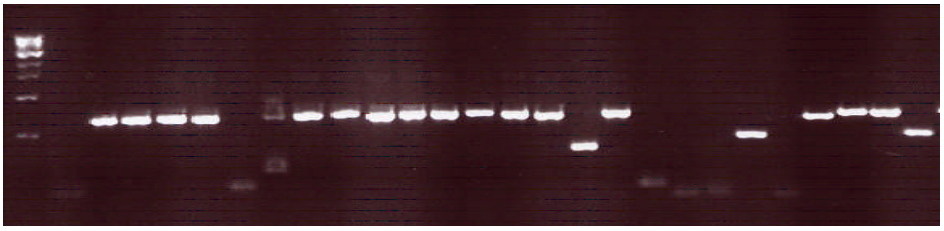


Figure 15. The spectrum of inaccurate DSB repairs identified by colony PCR using primers S_{11} and A_4 on maize samples.

.

5.3 Comparison Of Inaccurate Repair Events Among The Grasses Studied

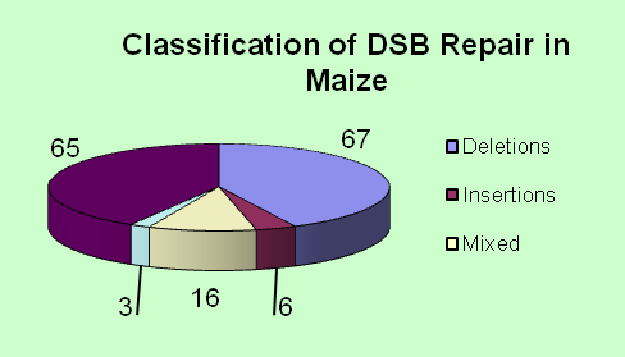
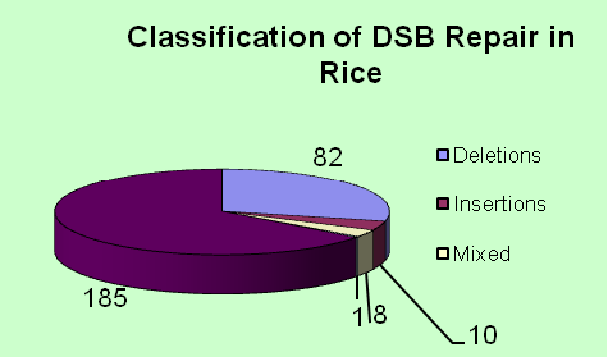
In order to characterize the outcomes of inaccurate DSB repair, the sequences at the induced DSB repair sites were obtained through the sequencing of 8 clones from each F_1 that contained both the I-*Sce* I ORF and its target site. These clones were derived from PCR across the target site, and the PCR product was then cloned into the TOPO TA vector without size selection of the PCR product. Species-specific difference and genomic location differences were also characterized.

Characterization of the I-*Sce* I cleavage site for each doubly hemizygous F_1

All of the obtained transgenic sequences were aligned with the uncut target sequence by the program Clustal W, and 4 types of inaccurate repair events were identified (Figures 23 through 29 in Appendix B). These were deletions, insertions, insertions associated with

deletions, and single nucleotide substitutions. The number of clones bearing each type of repair event was summarized for each species. The results in Table 9 indicate a wide range of events from a small number of F_1 plants analyzed (23 for maize and 47 for rice). Hence, there were usually several different inaccurate repairs from each F_1 plant, indicating that the inaccurate repair events were somatically sectored within individuals. The absence of inaccurate repairs detected in some resultant clones further indicated that either many accurate repairs were occurring in these lines, or that I-*Sce* I was only cutting the target site at some time late in development in most F_1. By excluding the clones that were redundant within a single individual F_1, I could then summarize the independent inaccurate repair events.

Table 9. Summary of the clone sequencing result in maize and rice

	
• 67 had deletions of 1 to 434 bp • 6 had insertions • 16 had both insertions and deletions • 3 had single nucleotide substitutions • 65 had no detected sequence change	• 82 had deletions of 1 to 493 bp • 10 had insertions • 8 had both insertions and deletions • 1 had a single nucleotide substitution • 185 had no detected sequence change

The independent inaccurate DSB repair events associated with simple deletions were summarized in Table 10 and 11. In both maize and rice, more than 50% of the deletion events were found being associated with micro-homology (1-4 bp) at the broken ends. In maize, 35/61 (57.4%) deletion events were associated with micro-homology (Table 10). In rice, 42/65 (64.6%) deletion events were associated with micro-homology (Table 11). It appears that the broken

ends formed in rice at DSB were easier to find a micro-homology sequence by random base-pairing according to a SSA-like model of NHEJ pathway.

Simple insertions were also identified in both species (Table 12 and Table 13). There were 5 events identified in maize, and 10 events identified in rice. Although rice had a higher frequency of small insertions than maize, the size increased and the frequency of insertions at the DSB repair site in both species are much less than the size deleted and the deletion frequency. This appears a common pattern of DSB repair and a strong evidence of inaccurate DSB repair as a genome size decreasing mechanism in maize and rice.

Table 10. Compilation of independent DSB repair events associated with simple deletions at the I-*Sce* I site in maize F_1

Event	Deletion (bp)	Homology (bp)	Event	Deletion (bp)	Homology (bp)
A11215-7_A11321_10_9_E02	1	0	A1131-8_A11216-8_11_4_D12	9	4
A11324-4_A123-2_7_5_E08	1	0	A11324-4_A1123-2_7_1_A08	9	4
A11324-4_A1123-2_25_5_E10	2	1	A11324-1_A11215_4_2_B05	9	4
A1138-4_A11215_7_2_B10	2	0	A11215-7_A11321_10_2_F01	9	4
A11324-1_A11215_11_8_H07	2	0	A11324-4_A1123-2_7_3_C08	10	2
A11324-4_A1123_10_3_C11	2	0	A1138-3_A11217_10_4_D06	10	3
A11324-4_A1123-2_7_8_H08	2	0	A11324-4_A1123_10_5_E11	11	3
A1131-3_A11216-8_12_1_C01	2	0	A1138-4_A11215_12_6_F11	12	1
A11324-4_A1123-2_21_3_C09	2	0	A11324-4_A1123-2_25	13	0
A11324-4_A1123-2_21_8_H09	2	1	A11324-4_A1123-2_25_4_D10	13	0
A11215-1_A11313_10	2	1	A1131-8_A11216-8_11_7_G12	15	1
A11217-7_A1139_16_8_F04	3	1	A11315-7_A11321_10_11_G02	15	1
A1131-3_A11216-8_12_8_D04	4	0	A1138-3_A11217_5_4_E05	17	1
A11324-4_A1123-2_25_8_H10	4	0	A1138-4_A11215_7_1_A10	17	4
A1138-4_A11215_7_5_E10	5	4	A1138-4_A11215_7_4_D10	17	1
A11324-4_A1123-2_21_2_B09	6	0	A11324-1_A11215_10_7_G06	18	2
A11217-7_A1139_16_2_E02	6	0	A1138-3_A11217_3_1_G01	20	4
A11324-1_A11215_4_1_A05	7	1	A11324-4_A1123-2_7_7_G08	20	4
A1138-3_A11217_3_4_G04	7	0	A1131-8_A11216-8_11_5_E12	20	4
A11215-1_A11313_10_2_B12	7	2	A1138-3_A11217_5_4_D05	26	0
A1138-4_A11215_12_7_G11	8	0	A11215-7_A11321_1_1_A01	28	1
A11324-4_A1123_10_6_F11	8	1	A11224-4_A1123_10_1_A11	34	0
A1138-4_A11215_7_7_G10	8	1	A11216-8_A1139_11	45	3
A11324-4_A1123-2_21_6_F09	8	0	A1131-3_A11216-8_12_5_D01	57	0
A1138-4_A11215_12_1_A11	8	0	A1131-8_A11216-8_11_8_H12	65	0
A11324-1_A11215_10_1_A06	8	0	A1131-8_A11216-8_11_2_B12	65	3
A11324-4_A1123_10_2_B11	9	4	A11321-10_A11215-10_3_C08	179	4
A1138-4_A11215_7_8_H10	9	4	A11215-7_A11321_10_3_G01	300	0
A1138-3_A11217_10_1_A06	9	4	A1138-3_A11217_5_3_C05	338	0
A1138-3_A11217_3_7_H03	9	4	A11215-1_A11313_10_5_E12	434	0
A1138-4_A11215_12_8_H11	9	4			

Table 11. Compilation of independent DSB repair events associated with simple deletions at the I-*Sce* I site in rice F_1

Event	Deletion (bp)	Homology (bp)	Event	Deletion (bp)	Homology (bp)
R3212-5_B3_E04_024	1	0	R329-4_D07_057	11	0
R3222-10-C05_043	2	0	R3220-16_C3_C11_091	11	2
R3222-16_C1_D10_074	2	0	R3220-16_C1_B10_078	17	0
R3225-1_D9_D10_074	2	0	R3220-16_C3_A11_095	17	4
R3213-1-1_C09_075	3	1	R3220-16_C1_A10/F10_080	17	4
R328-3_C04_028	3	1	R3215-2_A10_080	17	3
R3218-3_A11_095	4	0	R3218-3_B11_093	17	3
R32LR_G10_F12_086	4	1	R3220-16-c3_F11_085	17	0
R3218-3_C11_091	5	2	R3213-1-1_A09_079	18	0
R327-1_A7_B02_014	5	4	R328-3_F04_022	18	0
R3222-4_H04_018	5	0	R3212-5_E08_056	20	3
R327-1_A6_C01_011	7	0	R3220-16_D01_009	20	3
R3212-5_B04_B05_045	8	3	R328-3_H04_018	20	3
R3221-15-_C7_D01_009	8	3	R3220-16_C1_C10_076	20	1
R3225-1_D7_H09_065	8	3	R3212-5_B3_A04_032	20	1
R327-1_A08_A03_031	8	3	R3222-24_D4_E07	20	1
R3220-5_B11_C08_060	8	3	R3222-4_C8_A02/G02_016	22	1
R3221-15-1_G02_004	8	3	R3220-16_C3_B11_093	37	4
R3212-5_A08_064	8	3	R3212-5_B5_G06/E06/C06/F06	37	4
R3220-16_F01_005	8	3	R327-7_G03_019	38	1
R3222-10-H05_033	8	3	R3218-3_F11_085	38	0
R3232-2-3_H10_066	8	3	R328-3_B04_030	38	0
R329-14_A06_048	8	3	R328-3_A04_032	44	2
R3212-5_G08_052	9	0	R3232-2-3_G10_068	46	3
R3220-16_C1_E10_072	9	0	R3213-1-1_B09_077	109	3
R3221-15-1_E02_008	9	0	R3220-16_B12_F09_069	114	0
R3225-1_D9_A10_080	9	2	R3220-5_B09_B07_061	114	0
R32LR_G10_H12_082	10	1	R3213-1-1_D09_073	118	1
R3220-5_F12_086	10	0	R3213-1-1_F09_069	119	0
R3213-1-1_G09_067	10	1	R3220-16_A01_015	156	2
R3218-3_G11_083	10	1	R3222-10_E05_039	188	3
R3221-15-1_F02_006	10	1	R3225-1_D9_F10_070	493	0
R3221-15-1_A02_016	11	0			

Table 12. Compilation of independent DSB repair events associated with simple insertions at the I-*Sce* I site in maize F_1

Event	Insertion	Filler DNA Sequence
A11221-10_A11215_10_5	52	TTTGCCGAGTGCTTCAAACACTCGCCAAAGAAGCTTATTCCAGTAGTGACGA
A1138-4_A11215_12_4	12	ATTACCTATTAC (Copy upstream sequence)
A1138-3_A11217_10_5	1	T
A11324-4_A1123-2_25_1	1	T
A11324-1_A11215_3_5	1	T

Table 13. Compilation of independent DSB repair events associated with simple insertions at the I-*Sce* I site in rice F_1

Event	Insertion (bp)	Filler DNA Sequence
R3220-16_C1	1	A
R327-1_A8	1	T
R3222-10_D1	1	A
R3222-24_D03	2	TA
R3228-3_E04	1	T
R3215-2_F10	1	T
R3220-16_B01	1	T
R3213-1-1_H09	1	T
R3220-5_D12	1	T
R328-3_D04	5	ATAGG

I summarized the insertions associated with deletions, and deletions associated with SNPs in Table 14 and Table 15, and named them as mixed event. Again, in both species, the trend of size decreasing was much greater than the trend of size increasing. In maize, the total size deleted due to these mixed events was much greater than rice (2,231 bp vs. 797 bp). The higher frequency of mixed event identified in maize (13/82 in maize vs. 8/84 in rice) and its significant large contribution to the size decreased was a significant difference of DSB repair pattern between maize and rice.

Table 14. Compilation of independent DSB repair events associated with both deletions and insertions at the I-*Sce* I site in maize F_1

Event	Size Changes (bp)	Deletion (bp)	Insertion (bp)	Filler DNA Sequence
A1131-3_A11216-8_11_3	+5	-23	+28	TTATACTCGACGGATCCTAGGGCAATGA
A11220-5_A11316_8_4	-2	-2	+0	With SNP
A11215-1_A11313_10_3	-2	-2	+0	With SNP
A11324-4_A1123-2_25_2	-8	-8	+0	With SNP
A1131-3_A11216-8_11_2	-17	-37	+20	TTACCAACAACAACAAACAA
A11324-1_A11215_4_5	-18	-42	+24	AACAATTACCAACAATTACCGTCA (copy downstream sequences)
A11324-1_A11215_4_8	-30	-30	+0	1 bp microhomology; SNP
A11215-1_A11313_10_1	-236	-236	+0	TGTTATaggaagttcatttcatttggaga; micro-homology with 'CC', a part of the target site flanking sequence
A1131-3_A11216-8_12_6	-251	-251	+0	micro-homology
A1131-3_A11216-8_11_7	-384	-487	+103	ACAACAAACAACAAACAATTATTCATCTGCAGGACGGAAATTATCGAC AAACAACATTACAATTACAATTACATTTACAATTACAAACAATTATTAAC GC Insert 103 bp and with end microhomology
A11321-10_A11215-10_2	-423	-429	+6	AAGATTT
A1131-3_A11216-8_11_5	-437	-465	+28	GTTAATAAAATGAGCTCTTATACTCGCC
A11217-7_A1139_16_5	-428	-466	+38	GAGCTCTTATACTCGACGGTACCTATTACCCTGTAAAT

Table 15. Compilation of independent DSB repair events associated with both deletions and insertions at the I-*Sce* I site in rice F_1

Event	Size Changes (bp)	Deletion (bp)	Insertion (bp)	Filler DNA Sequence
R3212-5_B5	-24	-29	+5	TAATT
R3225-1_D7	-20	-31	+11	AATGAAATGAA
R3220-16_C3	-48	-64	+16	AGAATTAATTCTCGAG
R3220-16_B12	-428	-442	+14	AATCCCACTATCCT; 2 bp end-microhomology
R3220-16_H01	-346	-350	+4	CTTC
R3220-16_C3	-11	-7	+18	TAGGTACCGTCAGTGGAG
R3215-2_E10	+13	-11	+24	TTTACAACAATTACCAACAATTAC (copy downstream sequence)
R3225-1_D9_H10	+39	-5	+44	CCAATGACTCCCGCCGTATGAAAGCAATACTCTAACATAACAGGG 2 bp microhomology 'CA'

To further characterize the inaccurate repairs associated with insertions, I searched the filler sequences by BLAST to the available databases. This analysis can help determine whether these insertions are micro-homology dependent, what kind of sequences they were and where they came from in the genome.

Since most of the insertions were 1-2 bps, it made no sense to search for homology to these tiny sequences. I selected the larger inserted sequences (at least 12 bp) for BLAST analysis. The largest insertion identified so far was a 103 bp insert in the maize genome (Table 14).

In both maize and rice, the inserted sequences had end microhomology (1 - 4 bp) between the broken ends. Also, there were 2 events in maize (Table 12 and Table 14) and one event in rice (Table 15) had the inserted sequence information from the neighbor sequences flanking the break site. All the other insertions in maize did not detect any significant homology in the NCBI nr genome databases. For the rice insertions BLASTed, R3225-1_D9_H10 (Table 15) hit *Oryza sativa* (*japonica* cultivar-group) genomic DNA, a segment on chromosome 11 with the highest similarity. Features flanking this part of subject sequence were both hypothetical proteins. Interestingly, the target site insertion R32-25, from which this insertion was derived, was also mapped on chromosome 11 (see Chapter 4). This inserted segment also had a high similarity to

sequences annotated as part of NADH dehydrogenase subunit 1 and subunit 5. Multiple genomic sequences in rice bear this same similarity to the NADH dehydrogenase, such as in the rRNA-45S ribosomal RNA on chromosome 12, in hypothetical proteins on chromosome 10, 6, 3 and 1, and in other locations. Hence, this appears to be a tiny mobile fragment of DNA that had undergone multiple rounds of amplification/transposition during rice genome evolution.

Comparing the data obtained for the monocots maize, rice, and sorghum with the published data on dicots (tobacco and *Arabidopsis*)

By comparing the inaccurate DSB repair data from several species, I hope to discover whether monocots and dicots follow the same pattern in the relationship of inaccurate DSB repair to genome size variation. I separated the events into two classes, size increasing and size decreasing, in comparison to the original segment size before DSB repair. The ratio of size increasing to decreasing was calculated for each species studied to see if there was any relationship between DSB repair tendencies and genome size variation. A correlation analysis between deletion frequency and the size of deletion corresponding to different genome size was made and statistical significance was tested. Thus, I can provide evidence regarding the relationship between the DSB repair and genome size variation during evolution. By excluding the redundant clones, the number of independent inaccurate repair events was obtained for each species (Table 16).

Table 16. The percentage of independent repair events identified in maize and rice F_1

Number of Independent Repair Events	Maize	Rice
Deletion only	61	65
Insertion only	5	10
Deletion and insertion	13	8
SNP	3	1
Total	82	84

The percentage of the four types of inaccurate repair events identified in this experiment is shown in Figure 16.

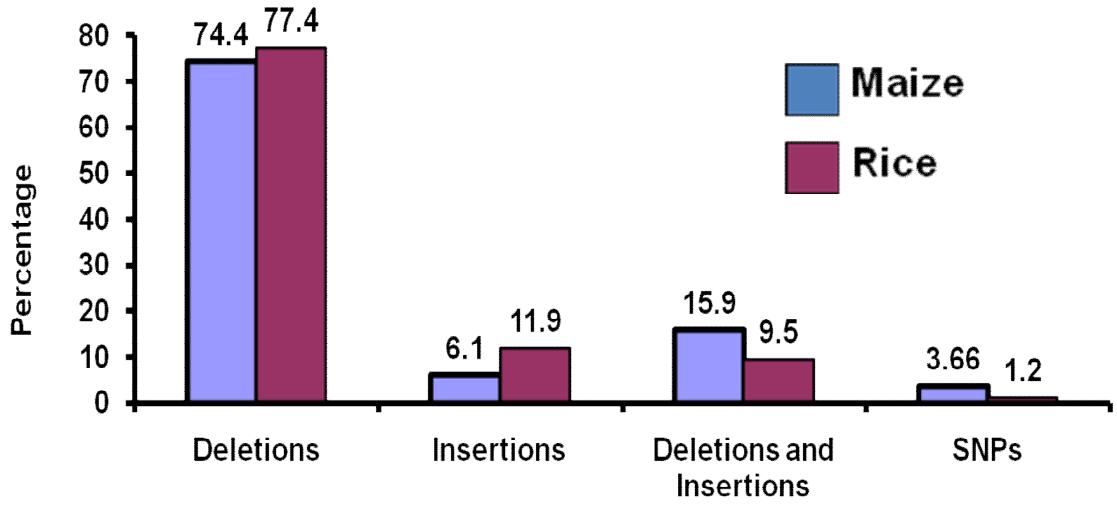


Figure 16. The percentages of the four types of inaccurate DSB repair outcomes were not found to be significantly different between maize and rice by χ^2-test ($0.05 < p < 0.1$).

In both maize and rice, deletions were found to be the most frequent inaccurate repair event (74.4% in maize and 77.4% in rice). Maize had a higher frequency of mixed (deletions plus insertions) events than rice (15.9% in maize vs. 9.5% in rice). Interestingly, in sorghum, a species more closely related to maize than rice, but without the ancient polyploidy event that happened in maize (GAUT *et al.* 2007; WALBOT and PETROV 2001), I did not identify such a high frequency of mixed repair event either among the 24 sorghum F_1 that were cloned and sequenced (Figure 29). It can be concluded that the smaller genome tends to have a simple repair pattern than the larger genome.

In order to see if the full spectra of inaccurate DSB repair events have any relationship with genome size variation, I summarize the detected size change events in Table 17. I did not find significant correlations between the size changes caused by the inaccurate repair and genome size variation by *Chi*-squared Test (Test Statistics $< \chi^2_{df = 2, \alpha = 0.1} = 4.605$). This means

that either there is no relationship, or that the relationship is too subtle to detect with a data set of this size. However, if one summarizes the actual nucleotide gain or loss across these 79 unique indels in maize and 83 unique indels in rice, one sees a preponderant net loss of DNA segment. The total DNA removed by these 79 indels in maize was 4,161 bp (the maximum deletion size was 434 bp), and the total DNA removed by these 83 indels in rice was 2,932 bp (the maximum deletion size was 493 bp). Clearly, the number of deletions at the repair site was not inversely proportional to their size in both species. No significant relationships were observed between the sizes of deletions and genome size variation between maize and rice by *Chi*-squared Test either (Table 17).

Table 17. Comparison of the effect of inaccurate DSB repairs on maize and rice genomes

Number of Inaccurate DSB Repair Events	Maize	Rice
Genome Size Decrease (Deletions)	73	71
Genome Size Increase (Insertions)	6	12
No change (SNPs)	3	1
Total	82	84

To further characterize the deletion repair events, the size distributions of genome size decrease events (including simple deletions and deletions associated with insertions) in each species were summarized, and are presented in Table 18. As one can see, small deletions ranging from 1-9 bp were the most common inaccurate repair event in both species. This agrees with the characteristics of haplotype diversity identified in rice (MA *et al.* 2004), as mentioned in Chapter 2. These results provide experimental evidence that the deletions induced by DSBs are the most significant feature of illegitimate repair and small deletions are more frequent than large deletions. Furthermore, this analysis did show that maize tend to have more small deletions, while the smaller rice genome had higher frequency of large deletions than maize (Figure 17).

Table 18. The distribution of deletion sizes identified in maize and rice F_1

	1-9 bp	10-20 bp	21-100 bp	>100 bp	Total
Maize	38	17	8	10	73
Rice	27	24	10	10	71
Total	65	41	18	20	144

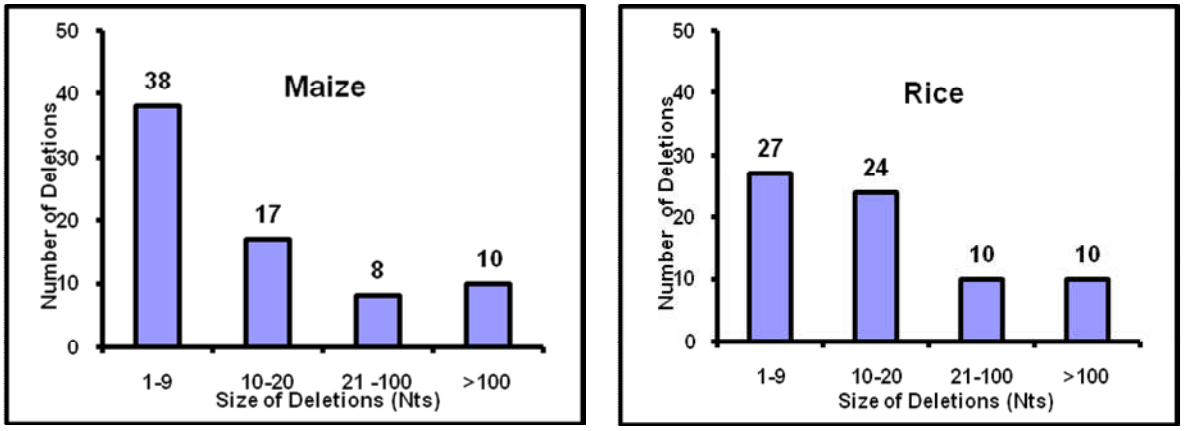


Figure 17. Size distributions of independent deletion events identified in maize and rice.

If we divide the sizes changing into two groups: 1-9 bp and >9 bp, such as Table 19, and test if the probabilities of small deletions of 1-9 bp is different between maize of rice, we can conclude that the probability of deletion size greater than 9 bp in rice was significant higher than that in maize at $\alpha = 0.05$ significance level. This conclusion supports the hypothesis that plants with small genome size tends to have higher frequency of large deletions, first proposed from observations in *Arabidopsis* and tobacco (KIRIK *et al.* 2000).

Table 19. Re-grouping deletion sizes identified in maize and rice F_1s for statistical analysis

	Size change 1-9 bp	Size change > 9 bp	Total
Maize (large genome size)	38	35	73
Rice (small genome size)	27	44	71
Total	65	69	144

Test Statistics = $(144 / 65*69* 73* 71)^{½} * (38*44-27*35) = 1.82$, $df = 1$, $Z_{\alpha 0.05} = 1.645$.

5.4 Identifying Unique DSB Repair Events Through Single-Seed-Descent

Because of the multiple different repair events observed in some individual plants, it was not possible to precisely quantitatively measure the relative frequencies of different classes of DNA repair outcomes. When several clones from a single plant yielded the same inaccurate DSB repair event, it was possible that they were separate clones of the same event or a rarer case of separate clones of the same outcome that happened twice or more independently. Moreover, because of the PCR approach that was taken, large deletions that removed at least one primer site or big insertions that made PCR less efficient would have been missed by this analysis. In order to rectify these problems, single-seed-descent was used so that only one event was recovered from each F_1 plant. This would allow more accurate quantization and also allow recovery of any type of rearrangement that was transmitted in that single descendent seed.

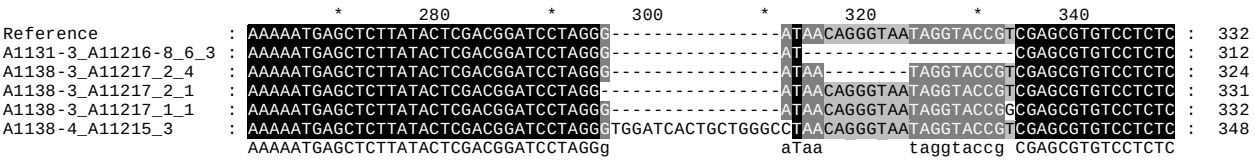
Maize F_1 with both transgenic constructs were used as female parent and crossed with B73 in order to get next generation seed. The comparable rice F_1 was self-crossed to get F_2 seeds. It was expected that 50% of resultant maize seed and rice seed would carry one unique inaccurate DSB repair event, while 25% of rice F_2 could contain two inaccurate DNA repair events, depending on the timing of the inaccurate repair during rice development. Only one inaccurate DNA repair event was sought per F_1 parental plant. I checked the progeny to see if any F_1 repair events were inherited, and what kind of repair event was inherited. Considering the segregation ratio, at least four progeny seeds descended from one F_1 individual were planted to make sure that at least one of them would having an inaccurate DSB repair sequence. For maize, the progeny would bear a single repair sequence derived from its F_1 female gamete. While for rice, due to the self cross, 25% of the F_2 should bear two repair sequences derived from each of the parent, and some would only have one repair sequences derived from only one of the parent.

A target site amplification were carried out after the isolation of the genomic DNA from these progenies, which were grown up till 2 weeks old for leaf tissue collection only. A total 79 maize progeny from 23 maize F_1 were obtained, and 40 (40/79, ~50%) were identified have the target site. A total of 119 rice F_2 from 18 rice F_1 were obtained and 89 (89/119, ~74.8%) were identified with target site. The segregation ratios agreed with my expectations.

Target site-specific PCR amplification products were purified using the Qiagen QiaQuick kit and sequenced finally. Forty maize progeny with target site amplification by primer pair of pTFBNEF-CTAACAGAACTCGCCGTGAAG and pTFBNER-GGTGTCCAGGTGAATATGTGG were sequenced and 44 of 89 rice F_2s with the target site amplification using the same primers were sequenced.

The results indicated that, although the progenies were carrying the target site sequence, most of them were carrying the integrated target sequences without any inaccurate repair event. This indicated that most of the repair events observed in somatic tissues were not transmitted, suggesting either some unknown gametic transmission problem or that the I-*Sce* I cutting did not commonly yield inaccurate repair events in the cells that eventually gave rise to the gametophyte. Whether this is due to more accurate DNA repair in this lineage of cells or to a lower activity of I-*Sce* I (perhaps due to specificities in *Ubi*-1 promoter expression) was not known. The sequences of the few identified inaccurate repair events are depicted in Figure 18.

Maize:



Rice:

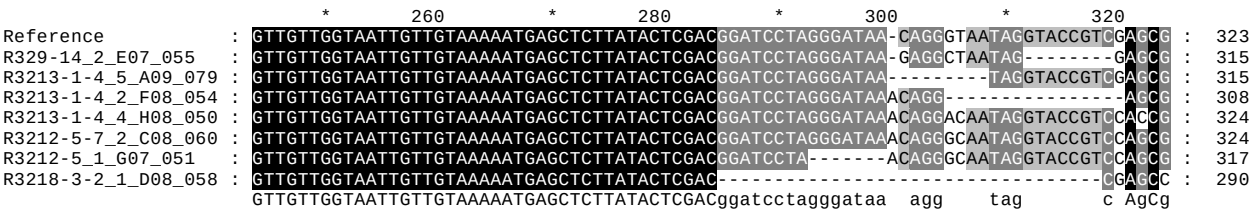


Figure 18. Inaccurate DSB repairs that were transmitted to progeny in maize and rice.

For maize, a total 5 different inaccurate repair event were identified in the F_1 maize progenies. A11138-3_A11217_2 had two progenies with different mutations. One progeny, A11138-3_A11217_2_4 had an 8 bp deletion, and the other progeny A11138-3_A11217_2 _1 had a 1 bp deletion at different position. In Rice, R3213-1-4 had 3 progenies identified with 3 different mutated target sites. Please note that there was an up to 25% chance of existing two different repair sequences in a single rice F_2. Also, due to the use of multiple pollen donors for generating rice F_1, the source of pollen donors was not recorded for rice study.

Identification of large deletions in single seed descent progeny

Since most our sequence characterization result were relied on PCR amplifications, for those complex repair events, such as large indels, or chromosomal translocations induced at the DSB site, which would be escaped from PCR or even more complicated assays like Southern Blot, it would be difficult to characterize them. However, it is these repair events that can contribute significantly to the genome size variations and structural changes, but not the small indels. Because of this, it is very important and necessary to find how frequent of these large indels events happened at DSBs. In order to do this, I used the single-seed-descent progenies derived from the F_1 individual with DSB repairs to identify those large indels.

The problem is that for those failed in the target site-specific PCR, we have no clue to know if there was a target site present or not in a segregated population. I need to know first which individual had the target site T-DNA integration but failed in the target site-specific PCR due to that large indels induced at DSB site eliminate the primer binding site. Two sets of *CodA*-specific primers (CodAFW1-CCGTAATCCTGCAAATTCAAC and CodARV1-GCTTTGGTCACGATGATGTCT, CodAFW2-TTTGTAGTCGATGGCTTCTGG and CodARV2-CGTTGAATTTGCAGGATTACG) were designed and used to confirm if there are target-site T-DNA presents in all the progenies of F_1 that I had planted (both maize and rice) and failed in their first round of target site-specific PCR amplification using primers of pTFBNEF and pTFBNER or S_{11} and A_4. If both the *CodA* amplifications have clear products, then I used the primer pair $S_1 + A_1$ to amplify the target sites, followed by primer pair $S_{10} + A_{10}$ for a second nested PCR amplification. These primers were used in Puchta's paper and they are further away from the target site than the S_{11} and A_4 primers. Their PCR products are about 3,000 bp for $S_1 + A_1$ in size and 2,300 bp for $S_{10} + A_{10}$ if no deletions happened. The $S_{10} + A_{10}$ PCR products are nested within the $S_1 + A_1$ products. I then directly sequenced $S_{10} + A_{10}$ PCR products after purification PCR products using the Qiagen QiaQuick kit.

Please note here that most or all small inaccurate repair events should have been amplified using the close pair of primers ($S_{11} + A_4$ or pTFBNEF + pTFBNER) during the first PCR screening. What I expected to miss were large insertions, large deletions or other large chromosomal rearrangements.

The results indicated that two large deletions were found in the single-seed-descent progenies, one large 1,746 bp deletion with insertion in maize (Figure 19) and one 1,127 bp deletion in rice (Figure 20). The event identified in maize turned out to be a mixed event. A 2

bp ('TA') micro-homology happened between the 5' broken end and the initial filler DNA sequence. Besides the large deletion size, the filler DNA segment has the first 18 nt homologous to a segment of reference sequence, and the following 15 nt of filler DNA was homologous to the 15 nt before the first homologous DNA segment in reference. The rest filler DNA was not found where they came from yet. The one event identified in rice was a simple large deletion (Figure 20).

From the single-seed-descent event in maize and rice, I can also give an estimate of the frequency of the PCR failed to amplify the target site in F_1 due to the limitations of PCR to amplify the large deletions or insertions at the DSB repair site. In maize, three lines of progeny of 23 F_1 did not give any target site – specific PCR products (13%). In rice, 1 lines of F_2 from 19 F_1 did not have any target site – specific PCR products either. Therefore, we may roughly conclude that about 13% inaccurate DSB repair site in maize are too large and beyond characterization in this study due to PCR limitations. Similarly, 1/19 (5.3%) inaccurate repair sequences failed to be characterized in rice in this study. These failure frequencies can also indicate how frequent the large indels induced by DSB in maize and rice. Not surprisingly, maize tended to have more large indels at DSB repair site than rice. This agrees with my sequencing result based on the PCR amplifications on F_1 plants.

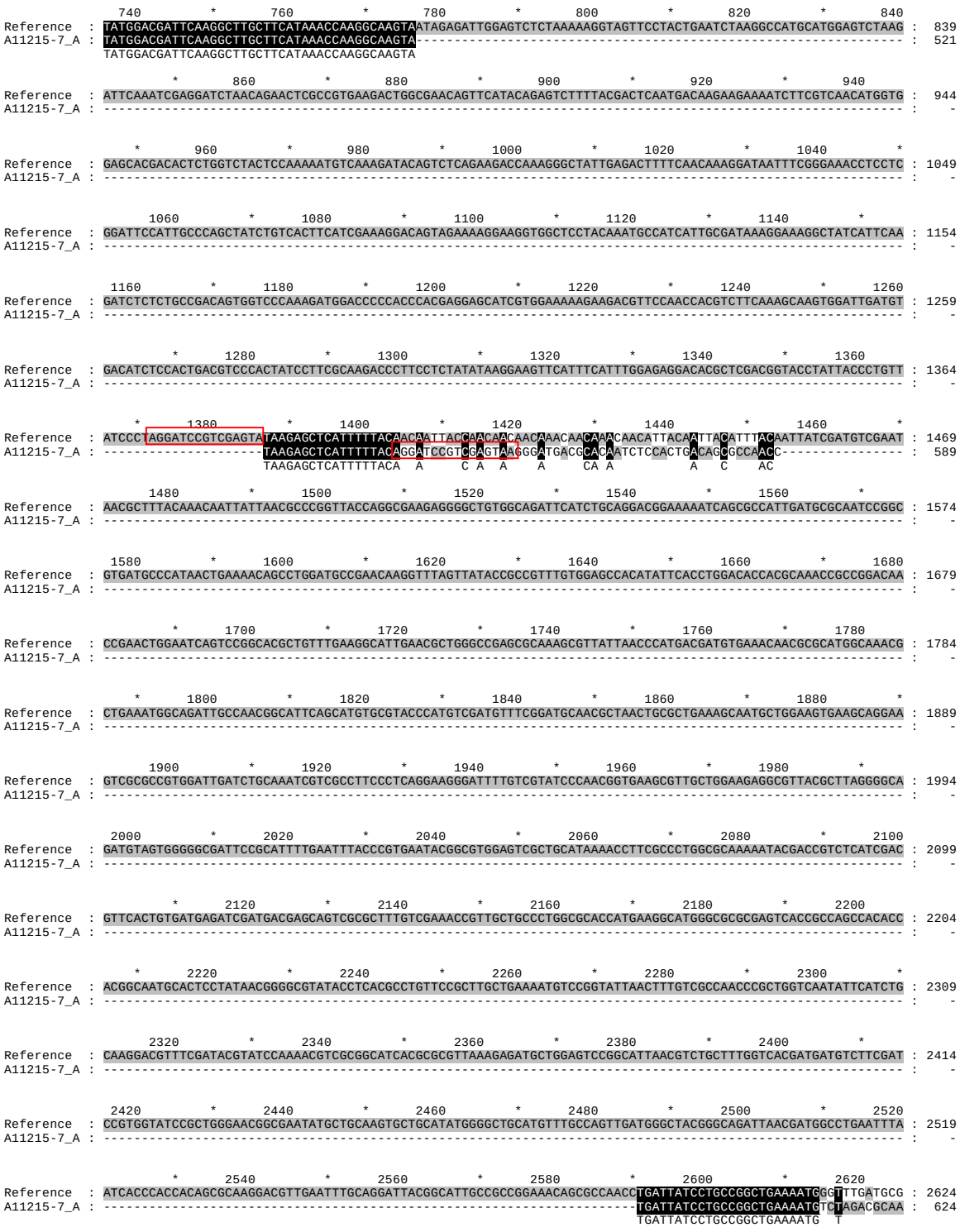


Figure 19. The large deletion identified in maize progeny. The two boxes indicate the homology between the filler DNA and the reference sequence.

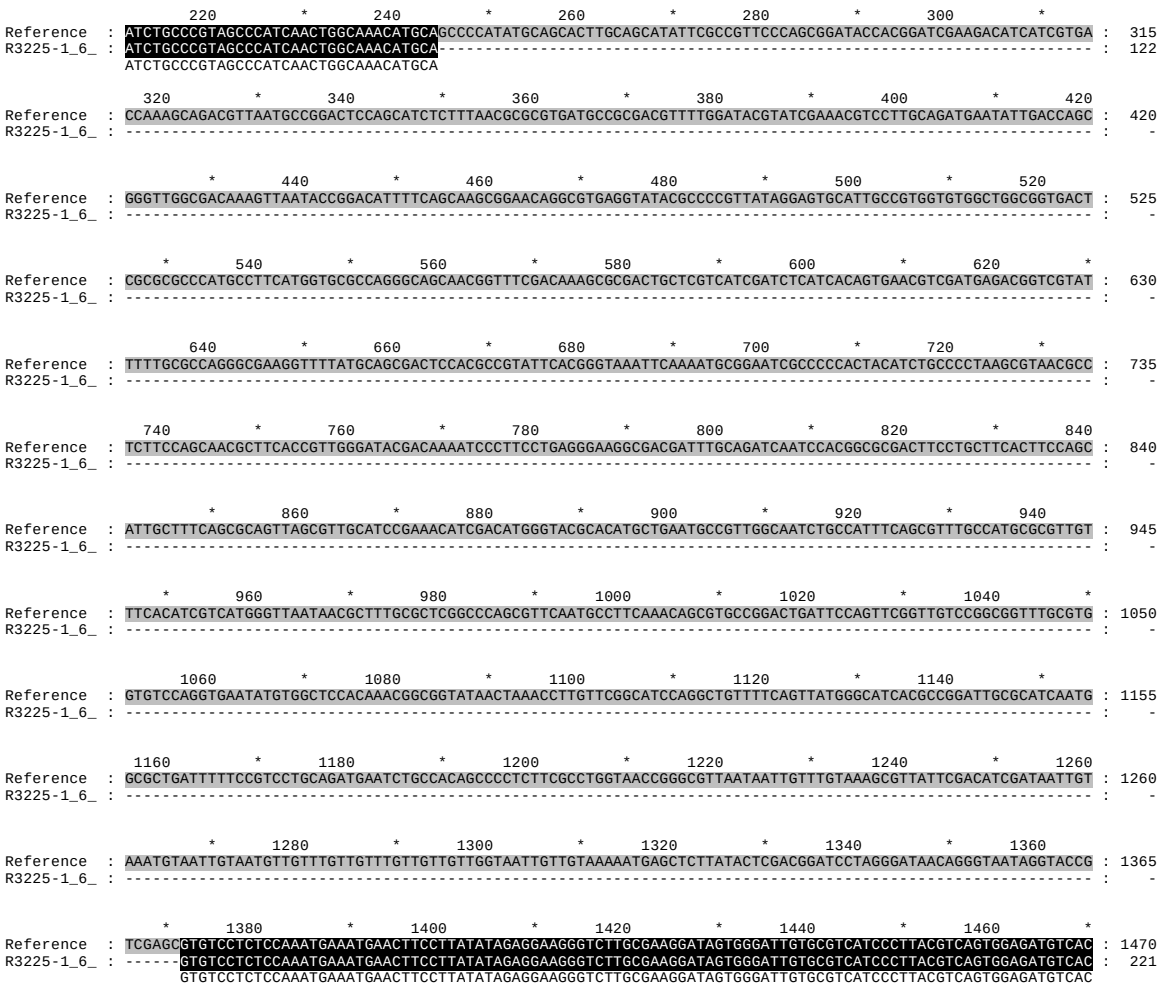


Figure 20. The large deletion identified in one rice F_2 individual.

CHAPTER 6

CLONAL ANALYSIS TO DETERMINE THE TIMING OF INACCURATE DNA REPAIR EVENTS AT I-*SCE* I SITES DURING MAIZE DEVELOPMENT

In order to determine the time of I-*Sce* I-induced inaccurate repair events, sector analysis was carried out on ears from the pollination of the hemizygous F_1 with B73 pollen. Unlike animals, plant germ cells differentiate from apical meristems after a long period of vegetative growth. An inaccurate repair event in the L2 cells of the shoot apical meristem can be transmitted to the next generation (IRISH and JENIK 2001). Six to eight seeds from several selected F_1 individuals with different target sites were planted and were pollinated by B73. For each of the ears harvested, one to three rows of seeds surrounding the cob were planted and tested for inaccurate DNA repair events. Each seed selected could represent a specific repair event that happened in F_1 gametophytic cells or could represent common clones of an event that happened earlier in development and would be inherited as a sector on the ear.

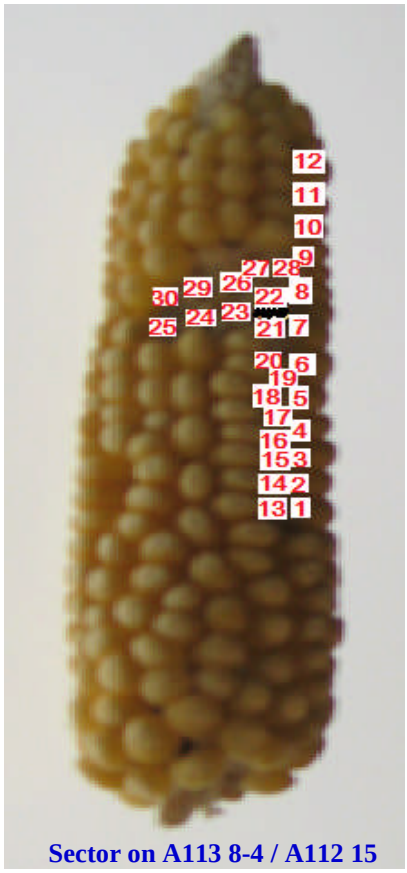


Figure 21. An example of a selected set of seeds for sector analysis.

From the frequency of different inaccurate repair events detected within one sector, I can determine how late in maize development an inaccurate repair event occurred. In addition, each event not in a sector (that is, each different event or identical identified from non-sectorial sites on the ear), will indicate the relative ratios of different kinds of rearrangements. Because the target site is hemizygous in the F_1, only half of the seed progeny from this cross are expected to contain the target site.

For three seed selections on three separate ears, the following results were obtained:

(1) **A113 8-4 / A112 15**: Fourteen of 30 progeny had the same 16-bp insertion at the target site.

(2) **A113 8-3 / A112 17**: Sixteen of 30 progeny had the same 1-bp deletion at the target site.

(3) **A113 24-4 / A112 3-2:** Thirteen of 30 progeny had the target site but no inaccurate repair.

From the presence of the same repair events identified within two randomly selected sectors on two separate ears, I can conclude that the I-*Sce* I induced cut and inaccurate repair events happened early in development, prior to the commitment to gametophytic development.

CHAPTER 7

DISCUSSION, CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS

7.1 Discussion

In this study, I-*Sce* I-specific repair events were induced and the specific inaccurate repair events (deletions, insertions, and insertions with deletions) were compared between maize and rice. Some data were also generated for sorghum. Extensive deletions were identified in both maize and rice. Rice tended to have a higher frequency and larger size deletions than maize. Also, the small genome size of rice and sorghum had a smaller number of insertions associated with deletions than the large genome of maize.

Comparing to Puchta's results (KIRIK *et al.* 2000) on *Arabidopsis* and tobacco, the deletion sizes identified in maize and rice (most less than 500 bp) averaged much smaller than what was identified on *Arabidopsis* (212 to 2,207 bp) and tobacco (222 to 2,294 bp). This may be because the primers I used (S_{11} and A_4) are much closer to the target site. But Kirik and coworkers (KIRIK *et al.* 2000) did not mention which PCR primers they used for their reaction I-*Sce* I site amplification. Because of the apparent size of their PCR products, I believe they used the S_1 and A_1 primers instead of S_{11} and A_4. Clearly, long distance PCR makes it more likely that one can identify larger deletions. The minimum deletion identified by Puchta lab (KIRIK *et al.* 2000) was over 200 bp. The reason they failed to identify < 200 bp deletions was due to excision of an amplification product of this size and smaller from an agarose gel after PCR. This

was necessary because, in their mixed transient transformation experiment in tissue culture cells, the great majority of the I-*Sce* I sites were probably not cut by I-*Sce* I, so they would have had a huge background of unchanged sequence at the site if they had taken all of the products of their PCR for analysis. In my study, we identified the inaccurate repair events by sequencing I-*Sce* I target sites because we knew that all had been in the same cells with an I-*Sce* I enzyme. Therefore, our sequencing results were more representative. In another study from Puchta lab (SALOMON and PUCHTA 1998), the deletions sizes identified in tobacco ranged from 88 to 1,322 bp. In order to compare my results across species in a more comprehensive manner, I believe a more extensive sequencing of long PCR amplification products across the target site will be necessary.

In tobacco, 40% of the repair events were associated with filler DNA (KIRIK *et al.* 2000) , while in my studies, the frequency of filler DNA was 22% (18/82) in maize, and 21% (18/84) in rice. The range of inserted DNA sizes in tobacco was also slightly larger in their studies (4 to 121 bp) than the insertion size I found in maize (1 to 103 bp).

Besides the above difference, Dr. Puchta's lab used a transient expression of I-*Sce* I to induce the cut, while I used a stable enzyme expressing line of I-*Sce* I. This difference in the experimental conditions should also be considered comparing these results.

7.2 Conclusions

- Inaccurate DSB repair often occurs in maize, rice and sorghum.

- Deletions occur more frequently than insertions at the inaccurate repair sites. Small deletions occur more frequently than large deletions. Deletion size is larger than

insertion size. In this small study, the two species with small genome sizes, rice and sorghum, tended to have less complex inaccurate repair outcomes than the species with large genome size, maize.

- The size distributions of deletion repair events in rice and maize are similar to those observed in dicot species. As observed for *Arabidopsis*, the small genome of rice tends to have more frequent large deletions compared with the large genome of maize. However, the size changes due to inaccurate repairs are not necessarily responsible for genome size variation in monocot species.

7.3 Limitations

I-*Sce* I expression in somatic plant cells

In our construct, I-*Sce* I expression is under the control of the maize *Ubi*-I promoter, which was reported to be a very strong and widely expressed promoter in grasses like maize (CHRISTENSEN and QUAIL 1996) monocots. Our results indicate that the enzyme is strongly expressed in the leaf in most of the transgenic plants. However, the transgenic gene I-*Sce* I expression is not guaranteed to be completely successful in all the plants or in all tissues during development.

Epigenetic modification, such as histone acetylation, methylation, or phosphorylation, as well as the recruitment of other ATP-dependent chromatin remodeling complexes could affect I-*Sce* I gene expression. Also, the I-*Sce* I sequence used in this study originated from yeast. It is hard to predict which modification will prevent I-*Sce* I expression in a specific individual. For

example, DNA methylation of the promoter region, or the I-*Sce* I ORF, or packing the I-*Sce* I structural gene into heterochromatin might inhibit the expression of the I-*Sce* I gene.

The genomic insertion location could also have an effect on I-*Sce* I gene expression. Even inserting near a gene might inhibit function though, if, for instance, the I-*Sce* I ORF is close to a 3' UTR of some strongly expressed genes, the expression of I-*Sce* I might be negatively subjected to the effect of an inversely orientated promoter by the creation of antisense RNA.

Limitations in using PCR to obtain the repair sequences at the break site

Clearly, PCR is an efficient method to identify sequences at the I-*Sce* I target site. It is much faster, less expensive and more sensitive than other techniques, such as Southern analysis. It must be pointed out that the two limiting factors of the PCR method to identify mutations at the cutting site would be the loss of the primer binding site by a large deletion or the inability to obtain a PCR product because of a very large insert. The first issue can be overcome by designing and using primers further away from the break site. Fortunately, the I-*Sce* I-induced small deletions in repair are much more frequent than the large deletions, and the pattern appears to be similar for insertions. Therefore, this should not be a big defect in this experiment.

Quantitative analysis of the inaccurate repair events

Because this study focuses on inaccurate DNA repair events in plant somatic cells, there might be many different independent inaccurate repair events happening even in one plant. The inaccurate repair events we identified by PCR followed by clone sequencing would represent the most frequent and developmentally earliest events in the leaf cells sampled. The earlier the I-*Sce* I expression, the more somatic cells developed later will possess the same repair event. If more cells will bear the same repaired sequence, the chance of identifying it in PCR reactions will increase. Those inaccurate repair events that happen later in development would have less

chance to be identified, depending on the number of colonies selected for sequencing. Our analysis of clonal sectors of inaccurate repair on the maize ear attempted to solve this problem, but uncovered only early events that did not appear to sector on the ears examined.

Based on our preliminary result, among the 2 X 96 clones we sequenced for maize, 34% (65 out of 192) of them did not show any polymorphism. In research in which the same enzyme was used on the human lymphoblastoid cells, the author identified 29 mutant clones among 929 sequenced, equal to 3.1% inaccurate repair events (HONMA *et al.* 2007). Three possible explanations were proposed. First, there were not cut because I-*Sce* I was not efficiently expressed in these cells. Since different somatic cells may have different physiological conditions, it is possible that the I-*Sce* I was not actively expressed in some cells. Second, the cells with DSBs entered into apoptosis instead of the repair pathway. Third, perfect joining happened and cannot be distinguished from the original recognition site. For these same reasons, we cannot measure the relative frequency of accurate versus inaccurate DNA repair events in our studies, although we are able to measure the timing of the events and the relative frequencies of different types of inaccurate repair.

7.4 Characterization Of Inaccurate Repair Events In *mre11, nbs1 and rad51* Mutants

Most homozygous *mre11*, *nbs1*, and *rad51* mutants in plants have clear defects in reproduction due to their important roles in meiotic recombination. The maize homozygous mutants in *mre11* and *rad51* are sterile, while the *nbs1* maize mutants are fertile. Each of these mutations has now been introduced into transgenic plants with an I-*Sce* I target site (C. Weil, Purdue Univ., pers. comm.). To characterize the inaccurate repair events in these mutants, first we need to introduce the target site into the heterozygous mutants by crossing the target site T-

DNA insertion line with the heterozygous mutants. Because the I-*Sce* I target site is also hemizygous, the probability of obtaining the ideal plants is 25%. The hemizygous mutant plants identified with *mre11*$^{-}$, *nbs1*$^{-}$, or the *rad51*$^{-}$ and carrying the target site are then crossed with the I-*Sce* I enzyme T-DNA insertion line in order to induce DSBs in the plants. The characterization of inaccurate repair events in these lines would be the same as the methods described above. The DNA repair mutations would be heterozygous at this point, and may have a dosage effect on DNA repair outcomes. We expect that the inaccurate DSB repair will be qualitatively different in these mutants compared with the wild type, perhaps with a different frequency of insertions versus deletions or different sizes of deletions.

Dr. Schnable's lab characterized the *MuDR*-induced double-strand break repair events in maize *rad51* mutant plants (LI *et al.* 2008). The *Mutator* transposon initiates a chromosomal DSB when it excises for transposition and is *subsequently* integrated at a new site through a "cut and paste" mechanism. Schnable and coworkers (LI *et al.* 2007) found that Rad51 plays a significant role in meiotic recombination, but not in somatic DSB repair in maize. The frequency of NHEJ events was increased in the *rad51* mutants that were defective in HR. The size change at the excision site in the *rad51* mutants ranged from -352 bp to +74 bp (LI *et al.* 2008).

In another study on RAD51 in diploid yeast cells, the researchers found that the presence of an active Rad51 protein inhibits the large inverted repeat-induced SSA pathway and proposed that RAD51 is to protect eukaryotes from genome rearrangement (DOWNING *et al.* 2008). These interesting results indicate that many more studies are needed to define the roles of DNA repair and recombination genes in the creation and/or prevention of genomic rearrangements.

REFERENCES

CHRISTENSEN, A. H., and P. H. QUAIL, 1996 Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res **5:** 213-218.

CORNEJO, M. J., D. LUTH, K. M. BLANKENSHIP, O. D. ANDERSON and A. E. BLECHL, 1993 Activity of a maize ubiquitin promoter in transgenic rice. Plant Mol Biol **23:** 567-581.

DOWNING, B., R. MORGAN, K. VANHULLE, A. DEEM and A. MALKOVA, 2008 Large inverted repeats in the vicinity of a single double-strand break strongly affect repair in yeast diploids lacking Rad51. Mutat Res **645:** 9-18.

FRAME, B. R., H. SHOU, R. K. CHIKWAMBA, Z. ZHANG, C. XIANG *et al.*, 2002 *Agrobacterium tumefaciens*-mediated transformation of maize embryos using a standard binary vector system. Plant Physiol **129:** 13-22.

GAUT, B. S., S. I. WRIGHT, C. RIZZON, J. DVORAK and L. K. ANDERSON, 2007 Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet **8:** 77-84.

HONMA, M., M. SAKURABA, T. KOIZUMI, Y. TAKASHIMA, H. SAKAMOTO *et al.*, 2007 Non-homologous end-joining for repairing I-*Sce* I-induced DNA double strand breaks in human cells. DNA Repair **6:** 781-788.

IRISH, V. F., and P. D. JENIK, 2001 Cell lineage, cell signaling and the control of plant morphogenesis. Curr Opin Genet Dev **11:** 424-430.

KIRIK, A., S. SALOMON and H. PUCHTA, 2000 Species-specific double-strand break repair and genome evolution in plants. Embo J **19:** 5562-5566.

LI, J., L. C. HARPER, I. GOLUBOVSKAYA, C. R. WANG, D. WEBER *et al.*, 2007 Functional analysis of maize RAD51 in meiosis and double-strand break repair. Genetics **176:** 1469-1482.

LI, J., T.-J. WEN and P. S. SCHNABLE, 2008 Role of RAD51 in the repair of *MuDR*-induced double-strand breaks in maize (*Zea mays L.*). Genetics **178:** 57-66.

LIU, Y. G., and N. HUANG, 1998 Efficient amplification of insert end sequences from bacterial artificial chromosome clones by thermal asymmetric interlaced PCR. Plant Mole Bio Reporter **16:** 175-181.

LIU, Y. G., N. MITSUKAWA, T. OOSUMI and R. F. WHITTIER, 1995 Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J **8:** 457-463.

LIU, Y. G., and R. F. WHITTIER, 1995 Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25:** 674-681.

MA, J., K. M. DEVOS and J. L. BENNETZEN, 2004 Analyses of *LTR*-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res **14:** 860-869.

PUCHTA, H., B. DUJON and B. HOHN, 1996 Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci U S A **93:** 5055-5060.

SALOMON, S., and H. PUCHTA, 1998 Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. Embo J **17:** 6086-6095.

SETTLES, A. M., S. LATSHAW and D. R. MCCARTY, 2004 Molecular analysis of high-copy insertion sites in maize. Nucleic Acids Res **32:** 1-12.

SHA, Y., S. LI, Z. PEI, L. LUO, Y. TIAN *et al.*, 2004 Generation and flanking sequence analysis of a rice T-DNA tagged population. Theor and Appl Genet **108:** 306-314.

WALBOT, V., and D. A. PETROV, 2001 Gene galaxies in the maize genome. Proc Natl Acad Sci U S A **98:** 8163-8164.

APPENDIX

A. THE SEQUENCES OF T-DNA CONSTRUCTS: PTFBNE AND PTFSCEI

>pTFBNE (sequence start from the right border)
AACAGCTATGACATGATTACGAATTCCAATCCCACAAAAATCTGAGCTTAACAGCACAGTTGCTCCTCTCAGAGCAGAATCGGGTATTCAACACCCTCATATCAACTACTACG
TTGTGTATAACGGTCCACATGCCGGTATATACGATGACTGGGGTTGTACAAAGGCGGCAACAAACGGCGTTCCCGGAGTTGCACACAAGAAATTTGCCACTATTACAGAGG
CAAGAGCAGCAGCTGACGCGTACACAACAAGTCAGCAAACAGACAGGTTGAACTTCATCCCCAAAGGAGAAGCTCAACTCAAGCCCAAGAGCTTTGCTAAGGCCCTAACAA
GCCCACCAAAGCAAAAAGCCCACTGGCTCACGCTAGGAACCAAAAGGCCCAGCAGTGATCCAGCCCCAAAAGAGATCTCCTTTGCCCCGGAGATTACAATGGACGATTTCC
TCTATCTTTACGATCTAGGAAGGAAGTTCGAAGGTGAAGGTGACGACACTATGTTCACCACTGATAATGAGAAGGTTAGCCTCTTCAATTTCAGAAAGAATGCTGACCCACAG
ATGGTTAGAGAGGCCTACGCAGCAGGTCTCATCAAGACGATCTACCCGAGTAACAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGATTCA
GGACTAATTGCATCAAGAACACAGAGAAAGACATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATT
GGAGTCTCTAAAAAGGTAGTTCCTACTGAATCTAAGGCCATGCATGGAGTCTAAGATTCAAATCGAGGATCTAACAGAACTCGCCGTGAAGACTGGCGAACAGTTCATACAG
AGTCTTTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACTCTGGTCTACTCCAAAAATGTCAAAGATACAGTCTCAGAAGACCAAAGGGCTAT
TGAGACTTTTCAACAAAGGATAATTTCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAGGTGGCTCCTACAAATGC
CATCATTGCGATAAAGGAAAGGCTATCATTCAAGATCTCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACC
ACGTCTTCAAAGCAAGTGGATTGATGTGACATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTT
GGAGAGGACACGCTCGACGGTACCTATTACCCTG↑TTAT↓CCCTAGGATCCGTCGAGTATAAGAGCTCATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTAC
AATTACATTTACAATTATCGATGTCGAATAACGCTTTACAAACAATTATTAACGCCCGGTTACCAGGCGAAGAGGGGCTGTGGCAGATTCATCTGCAGGACGGAAAAATCAGC
GCCATTGATGCGCAATCCGGCGTGATGCCCATAACTGAAAACAGCCTGGATGCCGAACAAGGTTTAGTTATACCGCCGTTTGTGGAGCCACATATTCACCTGGACACCACG
CAAACCGCCGGACAACCGAACTGGAATCAGTCCGGCACGCTGTTTGAAGGCATTGAACGCTGGGCCGAGCGCAAAGCGTTATTAACCCATGACGATGTGAAACAACGCGC
ATGGCAAACGCTGAAATGGCAGATTGCCAACGGCATTCAGCATGTGCGTACCCATGTCGATGTTTCGGATGCAACGCTAACTGCGCTGAAAGCAATGCTGGAAGTGAAGCA
GGAAGTCGCGCCGTGGATTGATCTGCAAATCGTCGCCTTCCCTCAGGAAGGGATTTTGTCGTATCCCAACGGTGAAGCGTTGCTGGAAGAGGCGTTACGCTTAGGGGCAG
ATGTAGTGGGGGCGATTCCGCATTTTGAATTTACCCGTGAATACGGCGTGGAGTCGCTGCATAAAACCTTCGCCCTGGCGCAAAAATACGACCGTCTCATCGACGTTCACTG
TGATGAGATCGATGACGAGCAGTCGCGCTTTGTCGAAACCGTTGCTGCCCTGGCGCACCATGAAGGCATGGGCGCGCGAGTCACCGCCAGCCACACCACGGCAATGCACT
CCTATAACGGGGCGTATACCTCACGCCTGTTCCGCTTGCTGAAAATGTCCGGTATTAACTTTGTCGCCAACCCGCTGGTCAATATTCATCTGCAAGGACGTTTCGATACGTAT
CCAAAACGTCGCGGCATCACGCGCGTTAAAGAGATGCTGGAGTCCGGCATTAACGTCTGCTTTGGTCACGATGATGTCTTCGATCCGTGGTATCCGCTGGGAACGGCGAAT
ATGCTGCAAGTGCTGCATATGGGGCTGCATGTTTGCCAGTTGATGGGCTACGGGCAGATTAACGATGGCCTGAATTTAATCACCCACCACAGCGCAAGGACGTTGAATTTG
CAGGATTACGGCATTGCCGCCGGAAACAGCGCCAACCTGATTATCCTGCCGGCTGAAAATGGGTTTGATGCGCTGCGCCGTCAGGTTCCGGTACGTTATTCGGTACGTGGC
GGCAAGGTGATTGCCAGCACACAACCGGCACAAACCACCGTATATCT

>Ubi-FRAGMENT1 _on the left border sice
AATAATGGAAAGGGCAAACCAAACCCTATGCAACGAAACAAAAAAAATCATGAAATCGATCCCGTCTGCGGAACGGCTAGAGCCATCCCAGGATTCCCCAAAGAGAAACACT
GGCAAGTTAGCAATCAGAACGTGTCTGACGTACAGGTCGCATCCGTGTACGAACGCTAGCAGCACGGATCTAACACAAACACGGATCTAACACAAACATGAACAGAAGTAG
AACTACCGGGCCCTAACCATGGACCGGAACGCCGATCTAGAGAAGGTAGAGAGGGGGGGGGGGGGAGGACGAGCGGCGTACCTTGAAGCGGAGGTGCCGACGGGTGG
ATTTGGGGGAGATCTGGTTGTGTGTGTGTGCGCTCCGAACAACACGAGGTTGGGGAAAGAGGGTGTGGAGGGGGTGTCTATTTATTACGGCGGGCGAGGAAGGGAAAGC
GAAGGAGCGGTGGGAAAGGAATCCCCCGTAGCTGCCGGTGCCGTGAGAGGAGGAGGAGGCCGCCTGCCGTGCCGGCTCACGTCTGCCGCTCCGCCACGCAATTTCTGG
ATGCCGACAGCGGAGCAAGTCCAACGGTGGAGCGGAACTCTCGAGAGGGGTCCAGAGGCAGCGACAGAGATGCCGTGCCGTCTGCTTCGCTTGGCCCGACGCGACGCT
GCTGGTTCGCTGGTTGGTGTCCGTTAGACTCGTCGACGGCGTTTAACAGGCTGGCATTATCTACTCGAAACAAGAAAAATGTTTCCTTAGTTTTTTTAATTTCTTAAAGGGTAT
TTGTTTAATTTTTAGTCACTTTATTTTATTCTATTTTATATCTAAATTATTAAATAAAAAAACTAAAATAGAGTTTTAGTTTTCTTAATTTAGAGGCTAAAATAGAATAAAATAGATGT
ACTAAAAAAAATTAGTCTATAAAAACCATTAACCCTAAACCCTAAATGGATGTACTAATAAAATGGATGAAGTATTATATAGGTGAAGCTATTTGCAAAAAAAAAGGAGAACACAT
GCACACTAAAAAGATAAAACTGTAGAGTCCTGTTGTCAAAATACTCAATTGTCCTTTAGACCATGTCTAACTGTTCATTTATATGATTCTCTAAAACACTGATATTATTGTAGTAC
TATAGATTATATTATTCGTAGAGTAAAGTTTAAATATATGTATAAAGATAGATAAACTGCACTTCAAACAAGTGTGACAAAAAAAATATGTGGTAATTTTTTATAACTTAGACATG
CAATGCTCATTATCTCTAGAGAGGGGCACGACCGGGTCACGCTGCACTGCAGGCATGCAAGCTTGCATGCCTGCAGGTCAACATGGTGGAGCACGACACACTTGTCTACTC
CAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTA
TTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCC
CACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATAT
CAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGT

>Ubi-FRAGMENT2_on the right border side
CAGGCCAGCGGCGGCGCTTCTTGTTCATGGCGTAATGTTCTCCGGTTCTAGTCGCAAGTATTCTACTTTATGCGACTAAAACACGCGACAAGAAAACGCCAGGAAAAGGGC
AGGGCGGCAGCCTGTCGCGTAACTTAGGACTTGTGCGACATGTCGTTTTCAGAAGACGGCTGCACTGAACGTCAGAAGCCGACTGCACTATAGCAGCGGAGGGGTTGGAT
CAAAGTACTTTAAAGTACTTTAAAGTACTTTAAAGTACTTTGATCCCGAGGGGAACCCTGTGGTTGGCATGCACATACAAATGGACGAACGGATAAACCTTTTCACGCCCTTTT
AAATATCCGATTATTCTAATAAACGCTCTTTTCTCTTAGGTTTACCCGCCAATATATCCTGTCAAACACTGATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCCAAGCT
CAAGCTCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAAT
GTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACAT
GATTACGAATTCTCATGTTTGACAGCTTATCATCGGATCTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTCTATCGCGTATTAA
ATGTATAATTGCGGGACTCTAATCATAAAAACCCATCTCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAATTCAACAGAAATTATATGATAATCATCGC
AAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTGCCAAATGTTTGAACGATCTGCAGGTCGACGGATCCAGTACTCGACTTATTATTTCAGGAAAGTTTCGGAGGAGA
TAGTGTTCGGCAGTTTGTACATCATCTGCGGGATCAGGTACGGTTTGATCAGGTTGTAGAAGATCAGGTAAGACATAGAATCGATGTAGATGATCGGTTTGTTTTTGTTGATT
TTTACGTAACAGTTCAGTTGGAATTTGTTACGCAGACCCTTAACCAGGTATTCTACTTCTTCGAAAGTGAAAGACTGGGTGTTCAGTACGATCGATTTGTTGGTAGAGTTTTTG
TTGTAATCCCATTTACCACCATCATCCATGAACCAGTATGCCAGAGACATCGGGGTCAGGTAGTTTTCAACCAGGTTGTTCGGGATGGTTTTTTTTGTTGTTAACGATGAACAG
GCTAGCCAGTTTGTTGAAAGCTTGGTGTTTGAAAGTCTGGGCGCCCCAGGTGATTACCAGGTTACCCAGGTGGTTAACACGTTCTTTTTTGTGCGGCGGGGACAGTACCCA
CTGATCGTACAGCAGACATACGTGGTCCATGTATGCTTTGTTTTTCCACTCGAACTGCATACAGTAGGTTTTACCTTCATCACGAGAACGGATGTAAGCATCACCCAGGATCA
GACCGATACCTGCTTCGAACTGTTCGATGTTCAGTTCGATCAGCTGGGATTTGTATTCTTTCAGCAGTTTAGAGTTCGGACCCAGGTTCATTACCTGGTTTTTTTTGATGTTTT
TCATCACTTCTTCTGTAGAGGTTTCTTCTTCCTCTTCTTTGTATTCTAGGTACAGTACTGGATCCTCTAGAGTCGACCTGCAGAAGTAACACCAAACAACAGGGTGAGCATCGA
CAAAAGAAACAGTACCAAGCAAATAAATAGCGTATGAAGGCAGGGCTAAAAAAATCCACATATAGCTGCTGCATATGCCATCATCCAAGTATATCAAGATCAAAATAATTATAA
AACATACTTGTTTATTATAATAGATAGGTACTCAAGGTTAGAGCATATGAATAGATGCTGCATATGCCATCATGTATATGCATCAGTAAAACCCACATCAACATGTATACCTATC
CTAGATCGATATTTCCATCCATCTTAAACTCGTAACTATGAAGATGTATGACACACACATACAGTTCCAAAATTAATAAATACACCAGGTAGTTTGAAACAGTATTCTACTCCGA
TCTAGAACGAATGAACGA

B. THE ALIGNMENT OF DIFFERENT TYPES OF REPAIR SEQUENCES

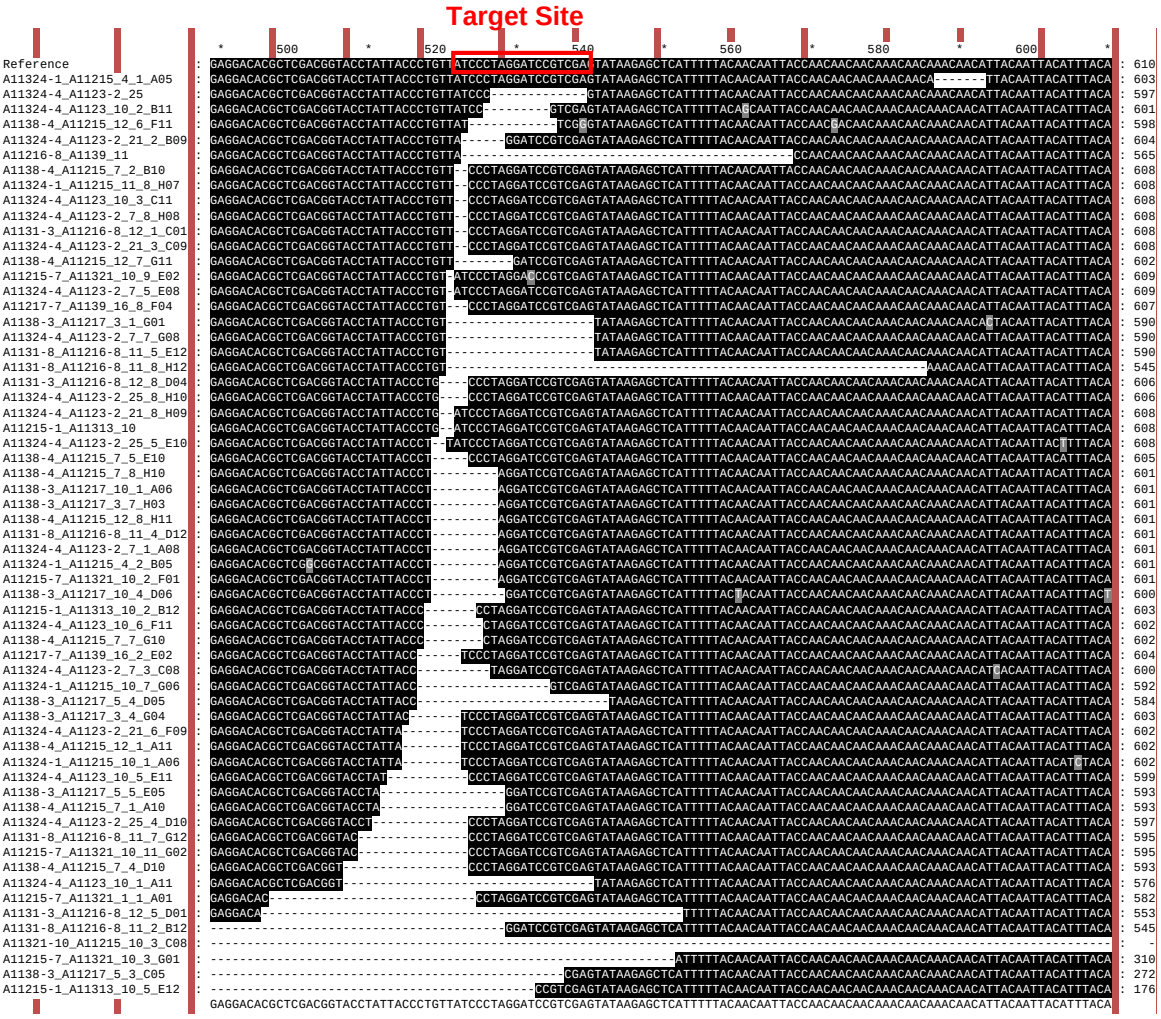
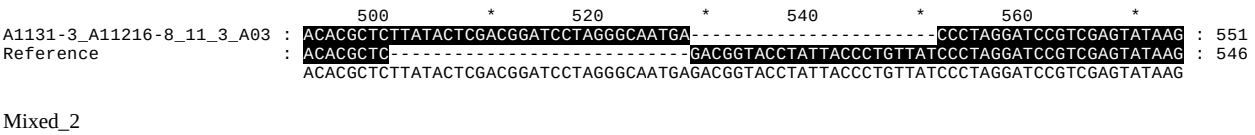


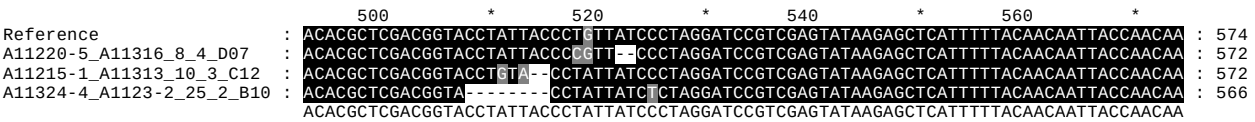
Figure 22. Deletions at the repair site identified in maize F_1 were aligned with the reference sequence. There are total 61 independent deletion repair events identified in maize.

Figure 23. Insertions at the repair site identified in maize were aligned with the reference sequence.

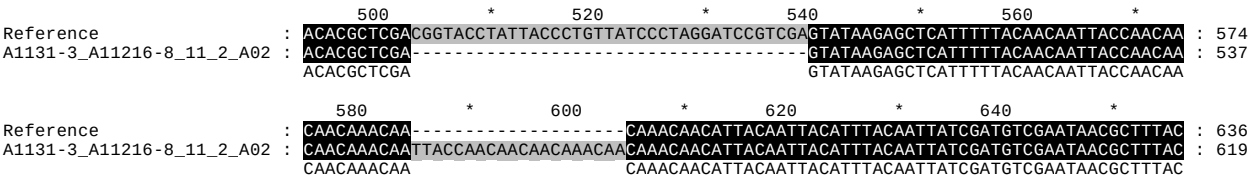
Mixed_1



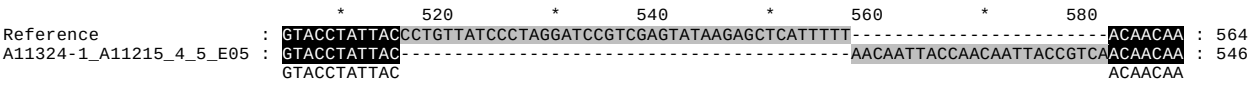
Mixed_2



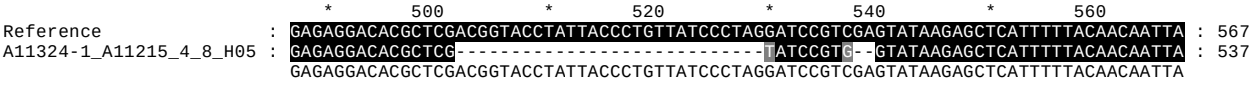
Mixed_3



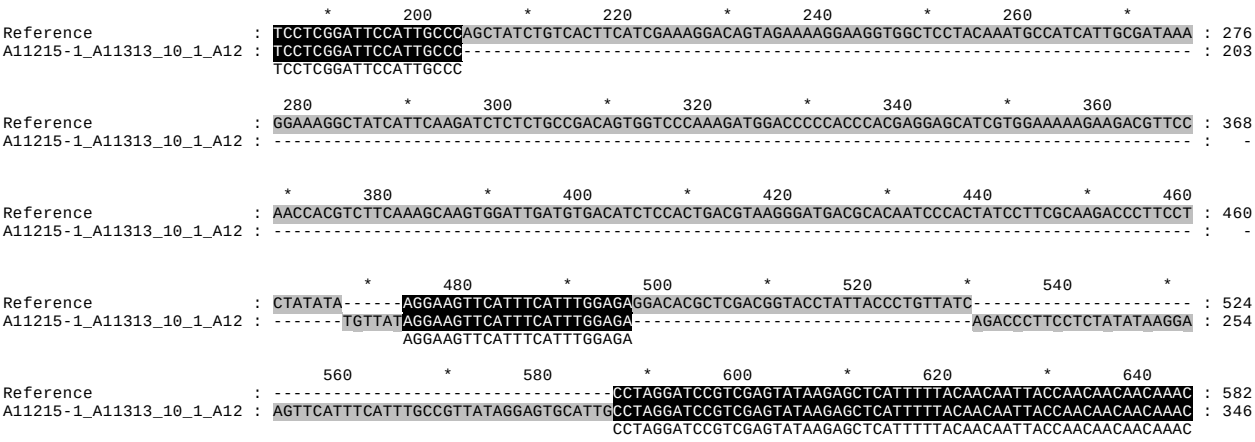
Mixed_4



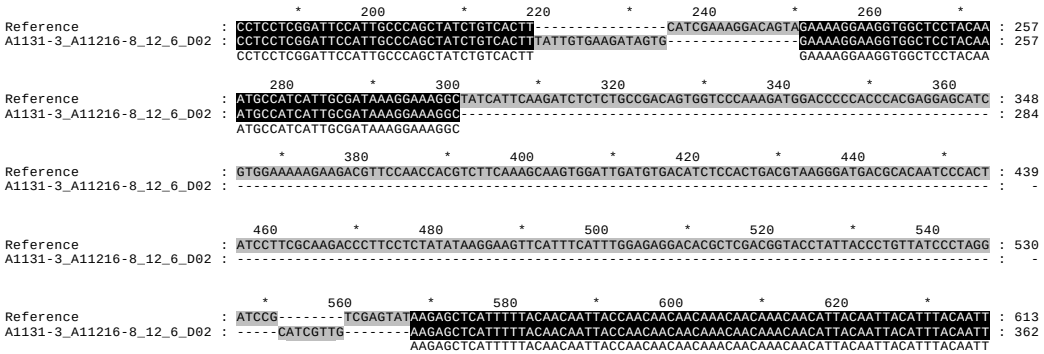
Mixed_5



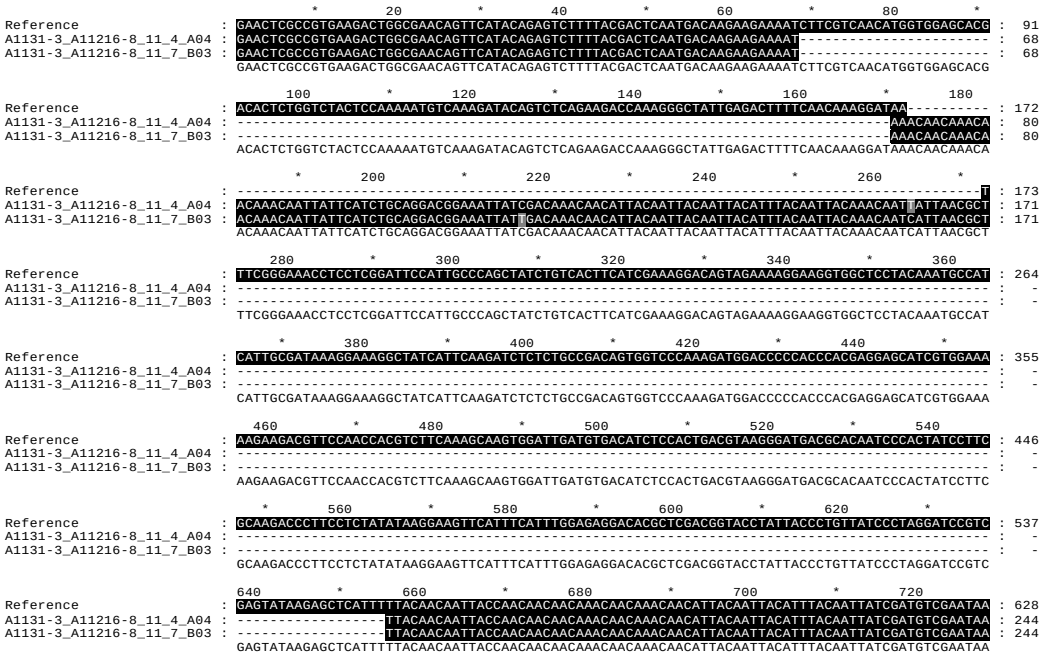
Mixed_6



Mixed_7



Mixed_8



Mixed_9

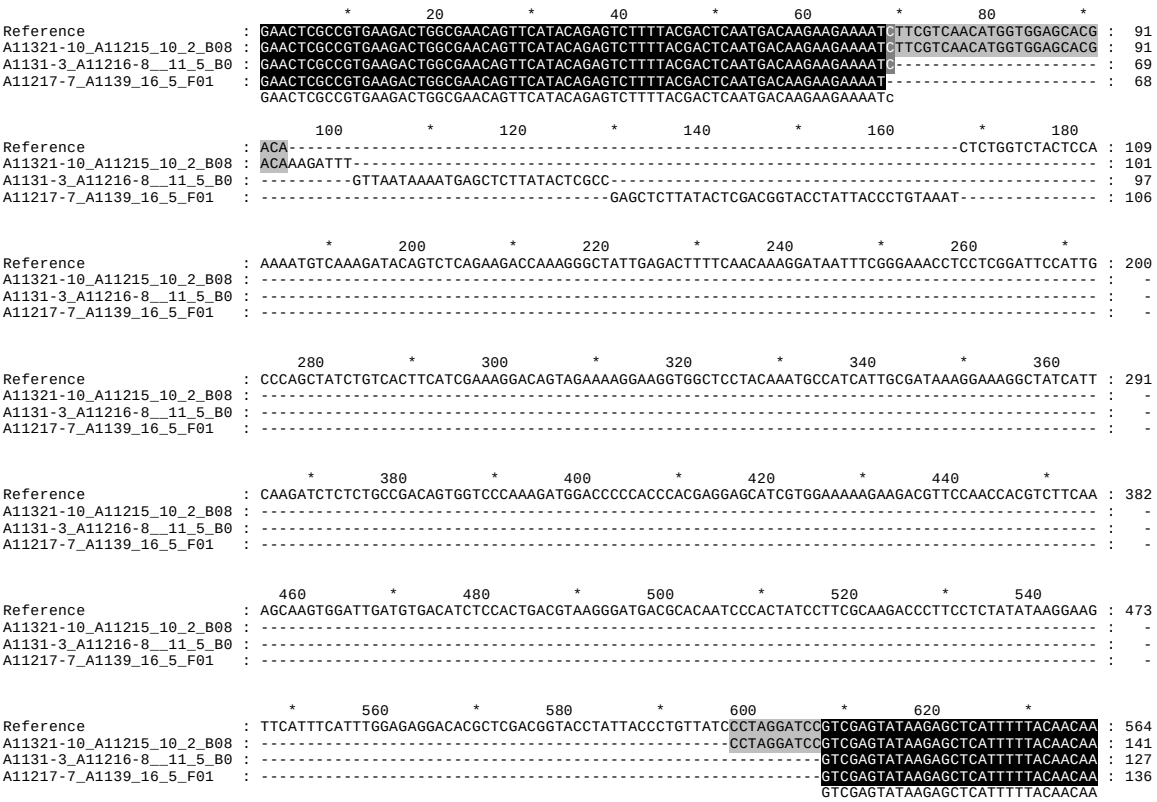


Figure 24. Both insertions and deletions (Mixed event) at the repair site identified in maize were aligned with the reference sequence.

Figure 25. SNPs at the repair site identified in maize were aligned with the reference sequence.

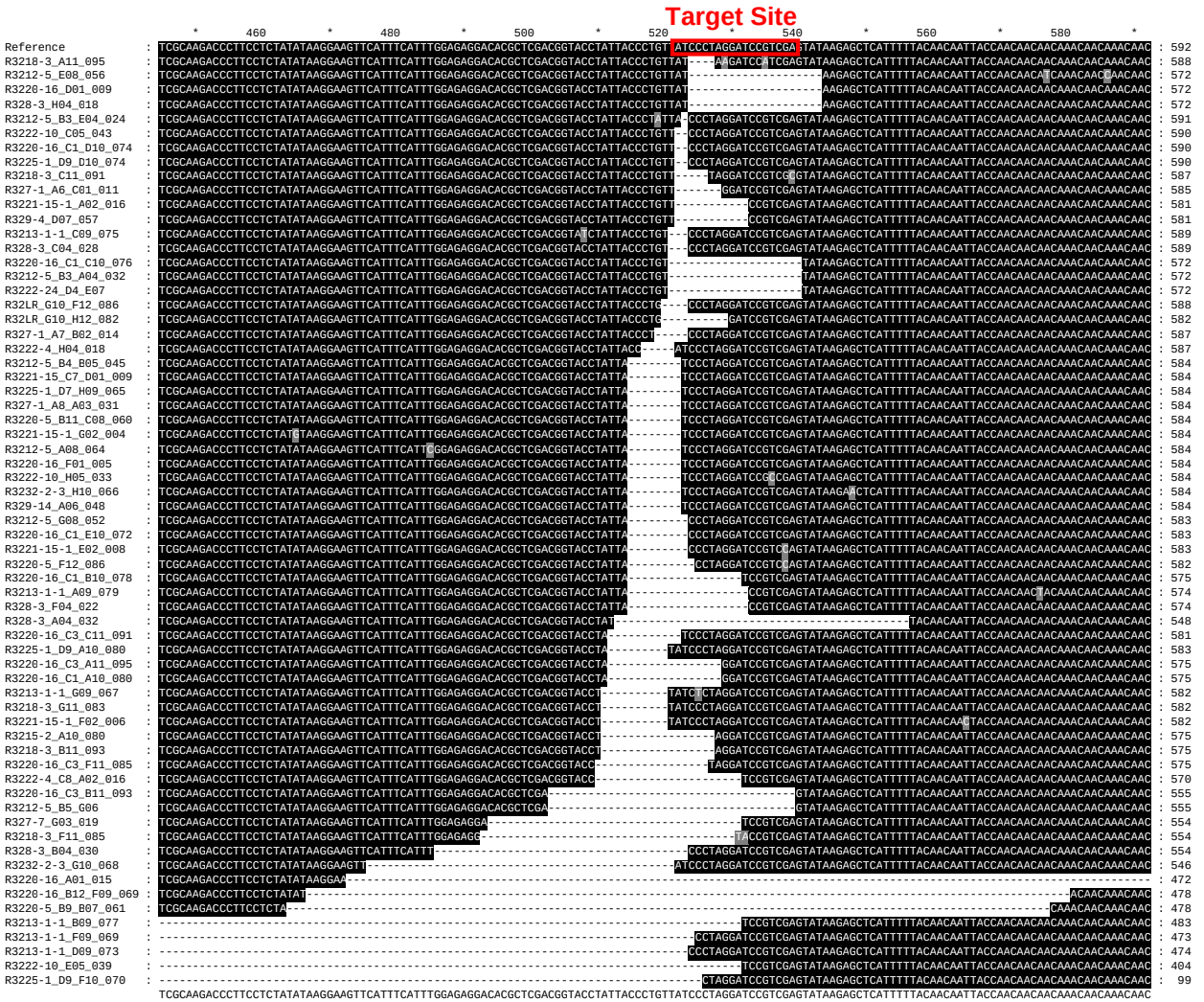


Figure 26. Deletions at the repair site identified in rice F_1 were aligned with the reference sequence.

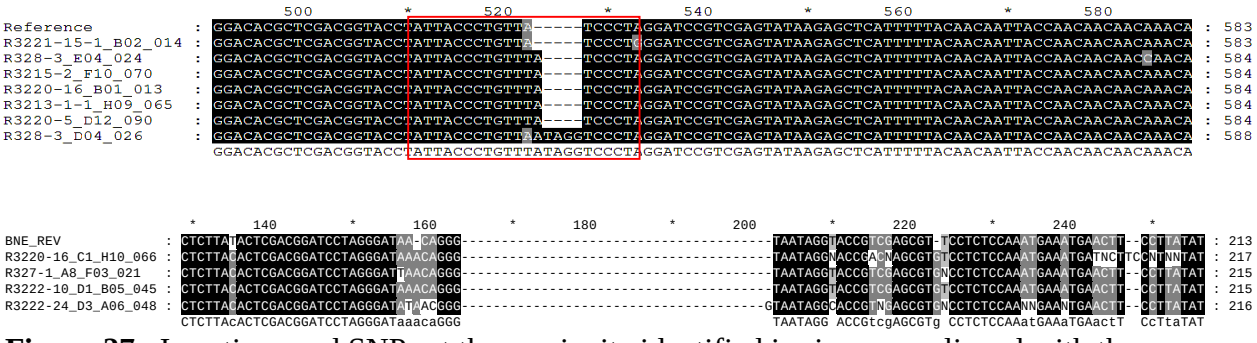
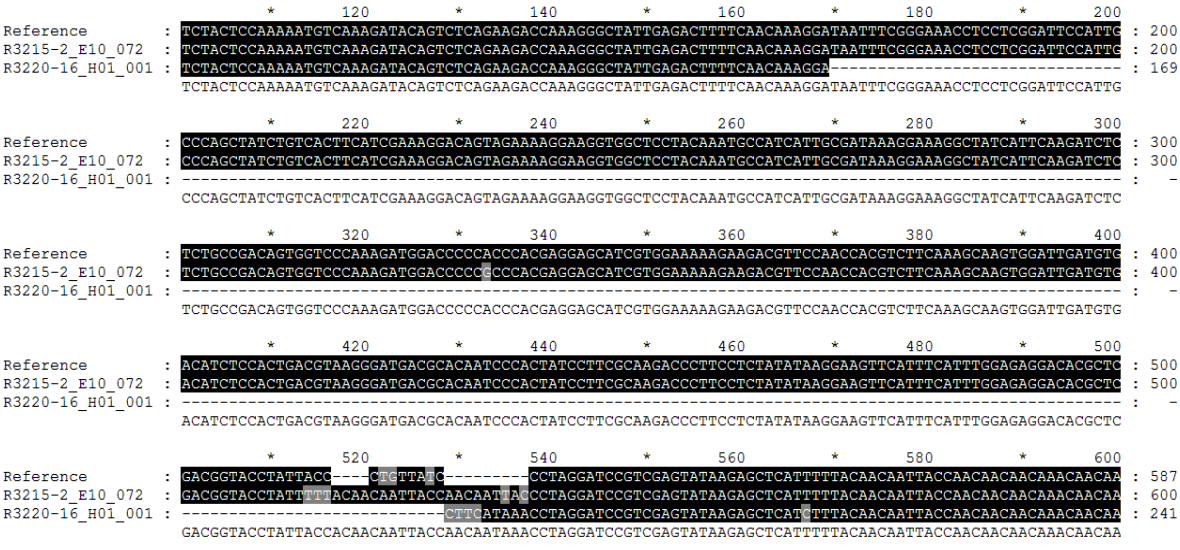
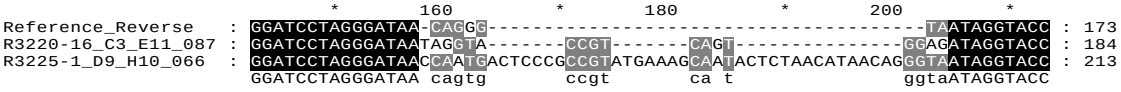


Figure 27. Insertions and SNPs at the repair site identified in rice were aligned with the reference sequence.

Mixed_1



Mixed_2



Mixed_3

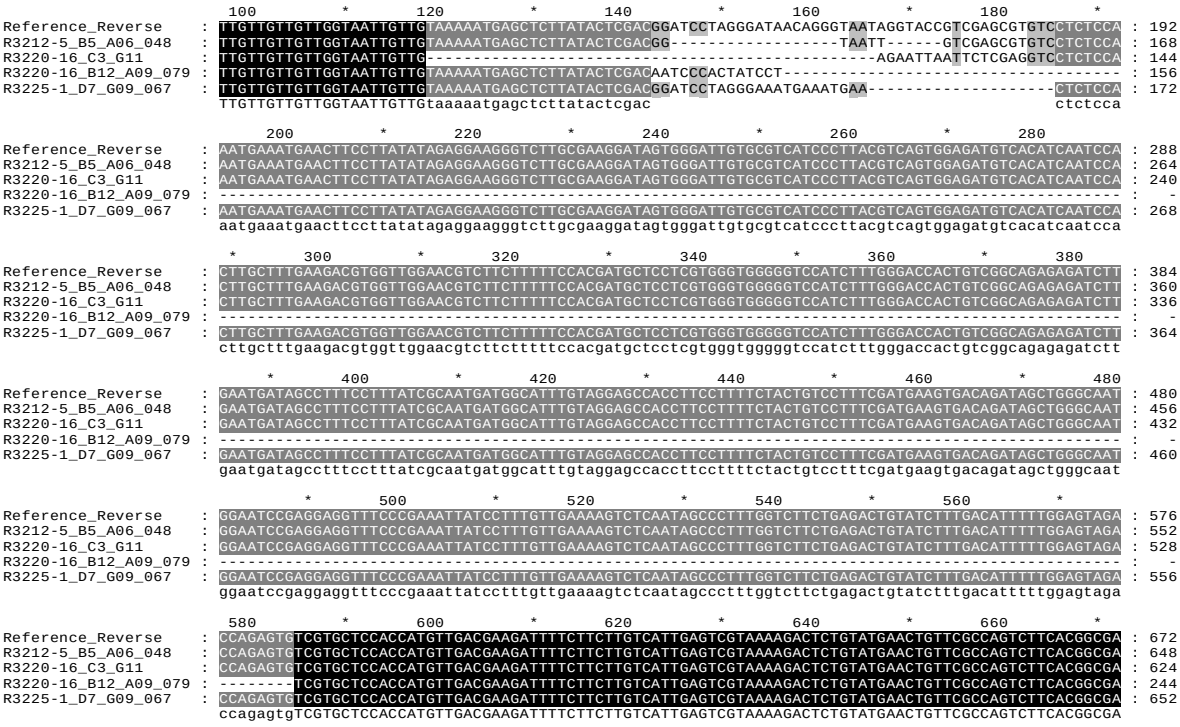


Figure 28. Both insertions and deletions (Mixed event) at the repair site identified in rice were aligned with the reference sequence.

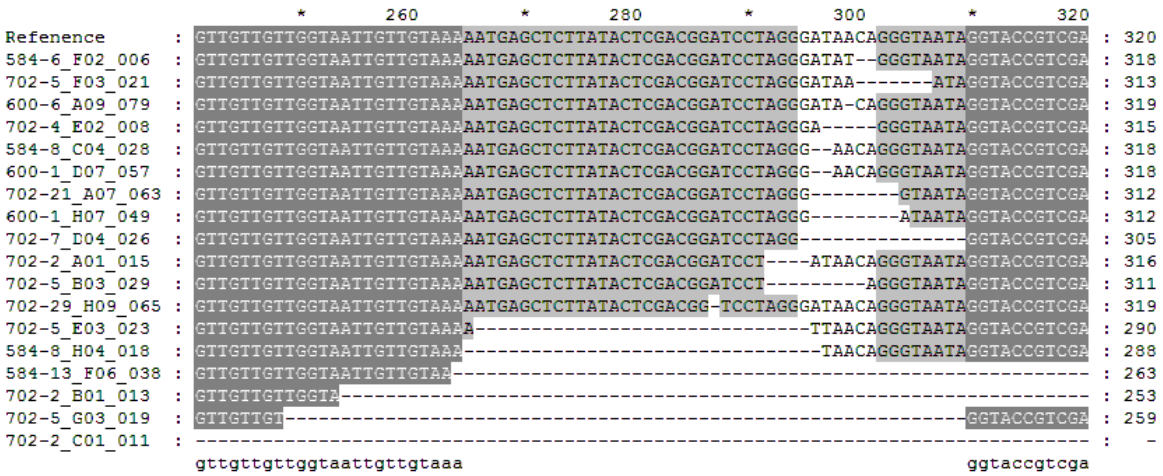


Figure 29. Deletion repair sequences identified in sorghum were aligned with the reference sequence. This was used as a species reference for maize and rice.

C. THE TARGET SITE T-DNA BORDER SEQUENCES THAT WERE USED FOR BLAST

The T-DNA border sequences in maize for BLAST

>A112-1_RB_AD1_D05
CCCGTAGACGTAGCAGTTTGTTTGCTTAACCCAATATTAGGCCATGTTCTTATTTAGTGAACCATGACCCCAAGCTGTTTTTTCCG
TTATCACTCCATGACCAAACACATAAAAATAAATGCTTAGATAATTTCCTTGTTAGACATAAAGTGTACCCGGGCTTAAAACGAGA
ATCTTGAGTTATATCTCCTTCCAAAACCGTGTCCTTGTTTGGTTGATGCAAGCTGATGTCCCTTATCCAGTTTTGCCTCTTCTTAA
GGATCATGTCGTTATTTTATCAACCTAACTTAGGAAAGGACTGCCCAATCTTTTTCTTGAATTTGATGCCTTTGGTTTCCTCATAT
CTATCAAAGGCATATCAACTTAATCTTATGATTATTTTCCCTCCGTATCTAGTCAGATACTAATCCCTGACCTTGTAATCTCTGCG
ATGGGCATAGAAATATGCTTGGACTGAAGTAACAGTCTCCCATCATACTCCTTTCATCAAGTTTAGGTTTTGGCTATATCGTGTTC
ATCTCACTCGACCCATTCAATTAAGAGGGATCCGCAAAAGTTGAGTAAGTATGTGCGATGAGCTCAGTCCAAACCTTTTCAAATCC
TCTAAGAAAACATTTTTACTAGATCATGAATGCTCTTGGAAAAAACACTTCCACTCCCCACTAAGATTTTCCTTTGGTTCACCTCA
GTTTTCTCAAAATCTGTTCAACAGTTTCAAAAATCGGTTCAACCGGACCTAAAACCGGTTCAACCGGTTTTTGCACTGTTCACCTC
TGTCTCTTTCTAACCTGCCAGTCTGCCAGTCAATCCTGTCAGAAAGTCGTGTGCAGATTTTTGGAAAACCGGTTCAACCGATTTTG
GCCTGGCTCAACCGGTTTTGGAGTCGGTCGAGTCTCCGGCTGAGTCTCCCAGTGAACCG

>A112-1_RB2_AD1_G01
CCAGCAGTCTAGTAAGAATTTTCTGACATTTCTAAAAATCAACTCATAGGGATTATATGTTCTACATCTGAAAAAATCAAAATAAA
AATCCAAGATGTGAAAAGATAAATGCTACATGAGATAGTGCTTCAACTCTTTTCAGCACTGATGTTCTTGTACTTCTAAAAAATCA
AATCAACTGTATTCAATCTTACAAATTTTACAGATAGTAATAACATTAGAAAAAATAATCTTCTTCCAAAACATATCGAATGTGAC
CATACCCATTCAGAAGCTCACGGGTGTGAAGCTGGTCCATGAATTGATATCATAGTGTATGATCATTTCTGATGTGGTAAAGGCCC
CCATTTAACCTACACACTAGACAGACAATGATCCATGGATGAGGTAGCCATGCCAACCTGATTCTCAACGGTTTCCTCAATGACAA
CTTGGCCCCAGCACCAACACCTGTA

>A112-3_RB2_W4_D06
CAATCGTCTATGTCGTTTCCGCCTTCGTTTAACTATCAGTGTTTGATAAATTGTGGTGTAAACAAATTGACGCTTAGACAACTTAA
TAACACATTGCGGACGTTTTTAATGTACTGAATTAACGCCGAATTGCTCTAGCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGG
TTTTCCCTTTATGCTTCCGGCTCGTATGA

>A112-3_RB_AD9_F09
GCGTCAATGTCGTTTCCGCCTTCAGTTTAACTATCAGTGTTTGATAACACATTGACGCTTAGACAACTTAATAACACATTGCGGAC
GTTTTTAATGTACTGAATTAACGCCGAATTGCTCTAGCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGG
CCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGA
CGTTGTAAAACGACGGCCAGTGCCAAGCTAATTCGCTTCAAGACGTGCTCAAATCACTATTTCCACACCCCTATATTTCTATTGCA
CTCCCTTTTAACTGTTTTTTATTACAAAAATGCCCTGGAAAATGCACTCCCTTTTTGTGTTTGTTTTTTTGTGAAACGATGTTGTC
AGGTAATTTATTTGTCAGTCTACTATGGTGGCCCATTATATTAATAGCAACTGTCGCTCCAATAGACGACGTCGATTTTCTGCATT
TGTTTAACCACGTGGATTTTATGACATTTTATGCTTCGGGCTCGTATGA

>A112-4_LB2_AD0_1_G11
ACTGGGCAGATAGTAATTAAAACGTACACGTCCCCTGCTTGCTTGTATCTTCATCACGACGCTCACATCTGCCTTATCTCCTATCG
CTGCCACGACCTCAAGCCGCCGAACACCATACTCGAGGCCTGCTCGAGTCGCTTTCAACTCATGGCCATGGAGTCCGCCTCAGCCT
CTGTCTTCCACTCGGTTGCTCCTCTGCTGGCTACGCAGAGCCTTCACTAGCCATGGCCGCGCCCTCTCTGGTACCGAGCTCGCCCA
TACCAAGCTCCCTCAGCGTCCCGTCCATGGTCCTCCTCCACTCGCCCTGGCCATGCTTGCCTGCCGCGCCACACTGAGTTTCTGGA
GTCTGGCACGGTGAACTGGGGAGAGAAGCAGCCAAGCAGCTCCCTCACGCCGCGGCTTCCTCCTTCAGCTCGGCGCTCATCTCCCT
CTCTGCTCGGACCGTGCCCAGCTCCGCGCGTCGCAGTTCGCCCCAGCTCGTTGTCCTGCCTCCTTTGCTCACCGCGTCGGCCACTC
GCTGATGCTCCAACTCATCAGTCGTTGTTGTCTCATCCGCGTCAGCACGCGACTATGGTCGTGTTCATCGAATTCGCCAACGCTTT
GTTGCTGATTCGACTGTCGTCGCTTCGCGTGTTGTCGAGCCCGTCGAGCCTCGTTCTTCTCTGCTCGCTCTCGTCGAACACCAAAC
TCGATAA

>A112-14-3_LB_W4_UPPER_H10
AACTGAGTATAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACATTAAAAACGCCCTT
CCCAATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCCAAGCTCAAGCTCCAATACGCAAACCGCCTCTCCCCGCGCGTTGG
CCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCAC
TCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGGCGAGTGACAAACAAATCCTAGAACGCCTAACTTTGTGGT
ACCTGTTCCCTCTGGCTAGATCTCCCGAACAAAAACGGGACGGAAACATGAAAATCAGCAGCTTCCAAAAGCATTGACACATTGAC
ACCTGGAACCGTCTGGACTCTGGCCAAAGGTTTCAATCATGTCTTGGCCGGGTCCTATAACAGCCGGACCGCCACCCACCTTGGTC
CTATCACTATTGTGTTTCATGTTTCAACGGAATGTCTCCCACTCTGATGTCTTCACTTTCCCCAACTACATTAACCAATA

>A112-15-1_RB_AD9_D04
TCGTCATGTCGTTTCCGCCTTCAGAGCACCTATAGGGGCTCGATGGCCAGCAGCAGGGGGGCTCATGGCTGGGCAATGTGCGATGA
CAGATCTAGACCTACGATCGATGGAGCAGCGCAAGACAAAATTGCACCAACGATATCAACGATGCATGATAATAGGCTAATCTGCT
AATATCATTTGTATGTAAAGAGCAACAATAGCATGTGCATAGAACATAAAAGATCATCATCGTGGCATTTAGAATTATGCGAGAAA
AATAACATACTATTTCTTACTATGAGATTAATCTAGTGCAAACCATCTATTATACTAGATCATTTGAAACTTCTACAATTATAGTG
TGTTTGGTTAGCAGCACAAACACCTAAAACGGTTGGAGCAAGCGGACTGAAGGCGCCCCGGTGATGGTGGAGTAAAACCTATACGT
ACGATTCAGTGTGTTATAGCAATATTCCTGCTTTTCTTAAACTTACCATAAGAAACAGCTAGTCCTCGTTCC

>A112-15-1_LB3_AD1_1_F10
CAATGGAACTATTAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGGCATGGGGGAGCAACCCGGTCCACGCGCTC
TTCACGTGGCCTCCACAGCGCCCGGCCCCGTCGTCGCACAGCCACAAGCTCCGCCGCATCTTTTTCTCTTCTCCGCCTTTCCGGTG
CCTTTAATTATTTTATTTATTTATTTCGCATATTGTTTTTTATTATTTTCTCGGCGAATAAGTACAACGACCCATGCACATCGTGT
GGCCAGTGTTTGCTTCTTCTCACTCGACCAATA

>A112-16_LB2_AD9_G06
ACTGGCGTATAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACTACCGCCGCCCATCA
TCCTCAAACGGCCGCGTAATAGATCGAGCCTCCCTCACTACACTGGTCGCCTGTTTATATATACAGCACCAACACCTGAAATGCAC
CAGGTGGGATTGTCCAAAAGGAGGCCAGCTAAATAGCTCACCGGAGTTTGGAATCGTAGGGGAGGATATGTGGACTCCATAATAAT
ACCAACCAACATATCAAGCACTCCAAATCTTATAATTACTAACCTATATACAATGTTGTGAAAATCACACCCAGGGAAAACACAGT
CAACAGTTTTCGGTCCTAATTTGCGCTTCTTGTTAATTGGCACATTCACTTTAGCCAAACAACCCCAGGTTTTTTTTATAGGAGAT
ATTTAATTTTCTCTTTTCCCATTCCTGGAAGGGAGTGATTTTTTTGTTCTTTGTGGAAACTTAATTCAGGATATGGAATGCCGTTA
AGA

>A112-18_LB2_AD1_G05
TCTGGGGTATTAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTCTAAGCGTCACCGCCGTAGACAT
GATCACGATGTTCCAGTGTTGGGATCTTCAACTGCAGGTTGGTGGGGGTGATCACCGTTGCTCGCCTCATGGGATTCATATCTTCA
CCAGGTGGGTTGGGCGGCAGGAGTCCAAGAAAGAGCTCACCGGAGTCTGGAATCGTAGTGGAGCATATGTGGACTCCATAATAGTA
CCTACCAACATATCAGGTACTCCAGATCTTATAATTAAGAACCTATATACAATGTTGTGAAAATCATACCTAAGGAAAACACAGTC
AACAGTTTTCGGTCCTAATTTGCGCTTCTTGTTAATTGGCACATTCACTTTAGCCAAACAACCCCAGGTGTGCAAATAGGAGATAT
TTAATTTTCTCTTTTCCCATTCCTCGAATGGTGTGATTTTTTTGTTCTTTGTGGGAACTTTATTCAGGATATGGCATGCCGTCAAG
ATCGTCTCACCCCACCATTCTGTAGATAGTTTTGAAGTCTCTAACATGGCGTTAACCAAATCAGTTAGAGTACGGTTCTTTCTCTC
AACAACCCCATTGGATTGTGGTGAGTATGGCGGTGTCCACTCATGAATAATACCATGTTCCACACAAAAATCAGAAAAATCACCTG
AAAAATATTCTCCACCACGATCAGACCTTAACCGTTTAGTTTTTCTCTCGAGTTGATTTTCCACTTCAGCTTTGTAGGTCTTCAAA
TAATGCAAAGCTTCGTCTTTTGACTTTAGTAGATACACATAATAAAATCTAGTAGAATCATCGATAAAAGTAATAAAATATCATTT
GCCTCCCTTAGTCAATATTCGTCGTCCCCCCCCCCACAAACAA

>A112-20-1_LB2_AD1_C10
ACTGGGCGTATAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACAACCATTAAAAAAG
GGTAAATTTTAGCGAAGGGGGCAGCTACAGTTGGCAGTTGGCGGGGAACTTGCTTTAGGAAATAATGAAAAGCGAGCACTCGAATC
TACACCCCACTGCTCTGTAAATTTGAGACTAACCACCAGACCTAACGTACCGTGTTTAGAAATAAACAATAATTATATTTGAATAA
ATATCTCAATACCTATGATATATAATAACCGTAGCAAAGCACATGCAACTGTCTAGTTTATATATAAGTGGTGTTTGGTTATGGGC
AAATATAGACAGTGGGAGGGAATATTATATCAGACACTGGATATCTGTCACCAAATTTTTTTACAAGTTGATCCATGTTTTCTAAT
CTTCACCGCTACTGAATAATTAAAGATTGTTAGTACTTCCTCCGTTCTTTTTTATTTGTCATGGTTTAGTTCAAACGCGAAAAACA
TCCGATAACAAAGATAGTATATGTTGGGGACTTGTTCTCAAATGCTATGAATCAAGAACAAGGCAACATAAAATATTAAATGTCAA
AGCCCTTCGTCCTTCGGATCATTATTTCCCTTCGGATATAATGAATCTAGGAGGACGAAGGTCATGAAAGACATGCCTTCATGATT
ATGATAAAAGATGAAGAACATTCATACAAAGTACAGAAAGTAACATAATTTTTTGTAAACATTATCATTAATCTATTTCCATTTAT
ATTATTTGTGCAAACAATAGTAGATTACAAATGTACCTTCGGCTTGAAAGGAGTGAGAATACAAGTGTAACACAAAAGCGAATGCC
AAGTCAGCGTGGACAGTACGGGAGTACTGTTCATCTATTTATAGGCACGGGAC

>A112-20_LB2_AD1_G08
CACTGGGCGATATCGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACAACCATTAAAAA
AGGGTAAATTTTAGCGAAGGGGGCAGCTACAGTTGGCAGTTGGCGGGGAACTTGCTTTAGGAAATAATGAAAAGCGAGCACTCGAA
TCTACACCCCACTGCTCTGTAAATTTGAGACTAACCACCAGACCTAACGTACCGTGTTTAGAAATAAACAATAATTATATTTGAAT
AAATATCTCAATACCTATGATATATAATAACCGTAGCAAAGCACATGCAACTGTCTAGTTTATATATAAGTGGTGTTTGGTTATGG
GCAAATATAGACAGTGGGAGGGAATATTATATCAGACACTGGATATCTGTCACCAAATTTTTTTACAAGTTGATCCATGTTTTCTA
ATCTTCACCGCTACTGAATAATTAAAGATTGTTAGTACTTCCTCCGTTCTTTTTTATTTGTCATGGTTTAGTTCAAACGCGAAAAA
CATCCGATAACAAAGATAGTATATGTTGGGGACTTGTTCTCAAATGCTATGAATCAAGAACAAGGCAACATAAAATATTAAATGTC
AAAGCCCTTCGTCCTTCGGATCATTATTTCCCTTCGGATATAATGAATCTAGGAGGACGAAGGTCATGAAAGACATGCCTTCATGA
TTATGATAAAAGATGAAGAACATTCATACAAAGTACAGAAAGTAACATAATTTTTTGTAAACATTATCATTAATCTATTTCCATTT
ATATTATTTGTGCAAACAATAGTAGATTACAAATGTACCTTCGGCTTGAAGGAGTGAGAATACAAGTGTAACACAAAAGCGAATGC
CAAGTCAGCGTGGACAGTACGGGAGTACTGTTCATCTAATTTATAGGCACGGGACG

The T-DNA border sequences in rice for BLAST

>R32_7-1_RB_AD1_D05
CCGTCAATGTCGTTTCCGCCTTCGTTTAACTATCAGTGTTTGAAGACGTTAATGCCGGACTCCAGCATCTTTTTAACGCGCGTGAT
GCCGCGACGTTTTGGATACGTATCGAAACGTCCTTGCAGATGAATATTGACCAGCGGGTTGGCGACAAAGTTAATACCGGACATTT
TCAGCAAGCGGAACAGGCGTGAGGTATACGCCCCGTTATAGGAGTGCATTGCCGTGGTGTGGCTGGCGGTGACTCGCGCGCCCATG
CCTTCATGGTGCGCCAGGGCAGCAACGGTTTCGACAAAGCGCGACTGCTCGTCATCGATCTCATCACAGTGAACGTCGATGAGACG
GTCGTATTTTTGCGCCAGGGCGAAGGTTTTATGCTTCGGCTCGTATAGATTGAGCAACTTATAATTTTTTTTTACTTCCCCAAAGC
CCTCCACAAATAAGCGGGGTGAGGTGTAGAGAGACACGAAA

> R32_7-12_RB_AD6_D06
CCGTTAAATGTCGTTTCCGCCTTCAGTTTAACTATCAGTGTTTGAAGACGTTAATGCCGGACTCCAGCATCTCTTTAACGCGCGTG
ATGCCGCGACGTTTTGGATACGTATCGAAACGTCCTTGCAGATGAATATTGACCAGCGGGTTGGCGACAAAGTTAATACCGGACAT
TTTCAGCAAGCGGAACAGGCGTGAGGTATACGCCCCGTTATAGGAGTGCATTGCCGTGGTGTGGCTGGCGGTGACTCGCGCGCCCA
TGCCTTCATGGTGCGCCAGGGCAGCAACGGTTTCGACAAAGCGCGACTGCTCGTCATCGATCTCATCACAGTGAACGTCGATGAGA
CGGTCGTATTTTTGCGCCAGGGCGAAGGTTTTATGCTTCGGCTCGTATTAACGACCATCTAACTGCTTTTTATTGCCAAATGCCCT
GCCCATGCATCGGTGTGGGGGGGGGTTGAAAAGCCAAACC

>R32_7_RB3_AD1_H02_002
TCGGNCGGNNGTTCCGCCTTCGTTAACTATCAGTGTTTANGNCTGNANTTCNNANNNTTACCANGATCCAGGCNAANCNTTNACGT
TAACANCNAGTAGCGCAATATATATGCTGTCGCTTGATCATGTTACAGCAGCATTGCTTGCATTGTGAGGTGGGGGTGATGCCACT
TTTTTGACAGCAAGTAAAGGAAGGAAACAGAGGCTCAAGAGAGATAGAGAAAATTAAAGTGGATGTGGTTGCTTGATTGTGCAACG
TACGTAAATAGGCAGGTACTCCATTTGAAGTTGAAGGCATTACCGCATTAGCTAGTCATTTGATTTGATGCGCTTTAAGGCCCTCG
CACGCAGGAGCATATATATAATTAAGCATGAATCGATGGATGCATGGATCCACGTACGGTTGACTCCGCGNGGTGGCGCTAGCTAG
CTTGATTGCTTGCTTAAAATCAAGTCACCGGAATGCGTGTGACACGCATTTCTTTCACCGGAATAAGCATAACGTCCGTCATTCGA
TCGAGTCGTCGTCCCGCTCCCGCATGGCGCGTGCGTACGGGTACACGTTGCGCGCGGCTTCATCATCCAGGCCTTTTGACCCTGCA
GACGCCGCGCGCCGACCACGACGTTGCACTCCCAATCACCTCCTGCCGCGTGCGTGCTCTATTCCTTTTCCTTCCTATCTCTATCC
TAACATGTGTCTGCCCCCAAATTCATCTCACTCGACCAA

>R32_8_LB4_AD1_E02_008
CAANCGCATGTGTATTAGTTGTCTAAGCGTCAATTTGTTTTTCCTTTCGTCTTGCTGTTGGTGCAAACTAGGGAGAAGGTTGCCCA
TACTTTTACCAATCATTTTCGTCAAGTGTTTCCCTAGAAGCAGCGCAGGCAGCGCACCAATGTTGTTTCGTATAACCGGCCAGCGA
AAGTAGTAGGTGAGAAATATGGGTCTCCTCTCTTACCAATATTCATCCAACCTACTAGTAACATACAACAAAGAGAGAAATTTCGG
ATAGGAATAAGTTCGTCTGAGGTCCCTTAACTTGTCAACGAATCCGATTTTCATCCTTCAACCGGAAAACCAGATACAACAGGTCC
CTCAACTATTAAAACCGGTGCAGATAAGGTCCCTCGGCGGTTTAGATGGCGGTTTTGGCTGACGTGGCGCCTACGTGGCAAATTCG
ACTCGGTCTTCATATGACATGGCACTTACGTGGCAATTTGATCCGAAAAAATAATAAACCCTGTGGGACCCACATGTCTGTTTCAC
ATACAAAATAATAATAAAAAAAATGGTGGGACCCACGTGGGCCCCACATGTCATGCTGCCATCCCCCCTTCATCCNACTCGACG

>R32_9_LB4_AD0_E05_039
TCANCGCATNTGTATTAGTTGTCTAGCGTCAATTTGTTTACACCACCATATATGCTGCNACCCCACCTGATNGGCTGTTGNGACAA
AGGGCGTGTGTAAGAAAGANTANANCTTTTACAGAATCGTAGCATTCNAGCNNNAGTTTTCTACNTTTTNCANNNNCGAATTCANN
GGATCTNGATGANNCTTGTNAAATGTTTGGAGAGAAGCAGCAGCTATCAAAAGCTCCNCAAANAGNCTCAGATCCTCTNCCTTCGA
AGNCNTCGTGCCTCTGCCAAATCCNTGNGCNGTGTTGAAGNCACATGAGTTCAAAAGGNATAGCGACTTTCAAGGNAGAGGATCCT
GATTCTTTACTTAGGCGTCTAGGCTGTTTACCATAGTTTCTCGCCATTGCAGTTTCTCTCCAAATTTTTAGCTTTNCNAATANNCT
ANTTTTTTTNTTANATTTTTCTNNNTNNACNANANNTNTGGANTNNNANNTTANNNNNCTNGNNNANNNNNNNAAANNCGGTNTTT
NTTTNAATANCAAAAN

> R32_12-2_RB_AD6_D10
GCGTTAAATGTCGTTTCCGCCTTCAGTTTAACTATCAGTGTCCACTACGGTAGAAGCCGGTCTTTATTAAGTGCAAATGAATTTGG
ATCATAACTCAACATCGACCTACGGTAGAAGCTGGTCTTTATTAAGTGCAAATGAATTTGGATCACAACTCAACATCGATGAGGAA
AATGACGTTTTTGGGTGGTTGATGATGAAATCAATCAATGGCGGTAATATAGGGTCCGATTTTGGCATGGTTTATCCCGTGAGGGA
TGAGACATCAGGGGAAGAAGACAATGTTTCTTTTTCTGTTGTGGACTCCACGAAACGATATTAAACTCTCAACCGATTTTGTAGGA
TCCCATCTCTCATCATCATGTTCCACACCATAAGTTAATTAGGTTTTTTTTTGTTTATTACAGTAAAGTCTATTTATTGTCTTTAT
ATTTTGCGCAATGTCCATTTTTCATATTGAACTGCAAAGCAATTATTTTTCACCCAAAACATTTAAAACCATAGCAAAGTACAACA
TTCTGAAGAAAACATTAGCTAGAATTGTGAAAAACATGAAAAAAACCCTACTCGTATTAACTGCCTCACCGTCACATCAATGCCAT
GGCAAAGACCAAATCAATATTGTCACGTAAGCGTCGCCGGACTTGGCTGGCCTGGCAGATAGGAAGGCTTTGGAGGGGTTGGGTTG
GGCTGGGTTGGAAAATGAAATGGCCCAAAGGAAATAGAGCACGCTGGGGTTAGATGAAGTTATTTTTCACAAGAAAAAGGATTAGA
TGGCTCCATCTCAGCCATAGAAAACAGTAGATGGCGGAGCTTTCGCAATCTACCTGGTGAGCCATGGCAAACTCCTGGCGATCCGG
ACTCACTGAATAGGGTATGATGACTACAGGCAAGTCGGAAATAGAAGT

> R32_13-1_RB_W4_D12
AACTGACTATGGTCGTTCCCGCCTTAGTTTAACTATCAGTGGTTTGAGATTAGAAAATAAACGAAGGGTTAGGATTAAATAGAAAA
GAGGGAATGAATGGTTAGAACTTAACGATGTCTTTTAGTAGGTGAAAAATTATTTCCTATTCCTATTAGCCAAACACTGCCTAGGG
AGTGACGGATTCGGATTAATAGAATTTGCGCATGGCCACTTAATAATAACGAGAGGAAACTACTGATCAAGTGGTTTCATCGAGGA
TAGTTATAACTCTAATATATTGTAATCATATGAATTAAGAAAAAGCGATTATCAGTGTTTCAAGGATGTAGGCTTCGACTAGACGC
CGTTCATAAATATCATGCTTAGTGTCTTTTATGCTTCGGCCTCGTATGATAGAGAATGTATTAAAGGTTTTCTTCTTTACAAAACA
ACATCACGTATAGATAAACATAGGTTCTAGAATTATGGTTTATAATATTTTTTGGAGAGTGGAACCAACATGTGATTGTCAAAATG
TGTGGGGGGGGGGGGATTTGGTAATTCTCGAGGGTTAATTAATTTTTTCCCGTTATTATATCCTATTATTATTATCG

> R32_14-1_LB_AD1_G12
ACTGGGGATATAGTAATTAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACTGCTCTTTACTATTTGATCG
AAGAGGCCAAATGATCCTTGCTTTCTTTTTCTTTGAGATCATTCCCGACTTGATGCGAAGGGTAGCTGTCTTTCCTTAGCCGCCGG
ACAGATAGTAAAAGTCAAGAGCTAGAGGTGATAGGTCGGATCCTTCCTCTTATTCAAGTCAATTGTGAATGCTGTGTACATACGTA
TAAACCACTTTTTTAATTCTTAATATCAGACAAGTAAAAACAATTGAACTTTCATTGAGTAAGTGTTATCAAGTTTTCCTCAAAAA
AAAAAACCTTGGAAAAAAAGGTGGGTTTGGGTTTCGGTTTGGAGGGGGAAGGGTTGAACCTCCACCTTTGGCGGGGCCGGAACCCC
TTAACAAAAAGGAAAAACTGGCAAAGGGGAAACACCAAACCATAAAAAAAAATTGTGGAATTTGGAATTGGGAAGGCCACCGTTTAA
AAATCTTCCGAACTGAAAAAAAAAAAAAACCCCCCGTTAAAATCCAGGGCTGGCAAGGCCCTGGCTTAAATTTTGGCTTTACTCCCC
CCCCCCCCA

> R32_14-1_RB_AD0_E02
TAGGGGACTATGGTCGTTCCCGCCTTTCGTTTAACTATCAGTGTTTGATTGTGAAACAATATTTCCAGTTCTCTTTCATGAATGGT
CATAGCAACTAATTTCTCTAGTTTGTTGGCAATGCTTAAATTGCCAGAATAATCCAGATAGAAGCATCCAATTTCATCTCAAAGCA
AATGTCTCTGTTCAATCATAATATTAATTTGTATTTACTATGAGAACAAATATTGCTCAATGCTAAGTTTTATTGCAACTTATGTC
ACTTGTGTTCACTCACTCGACGAATGCCAGCAACGGTTTCGACAAAACGCGACTGCTCGTCATCGATCTCATCACAGGGAACGTCG
ATGAGACGGTCGTATTTTTGCGCCAGGGCGAAGGTTTTATGCTTCCCCCGCCGTATGAATGTTATTTTGCGTTTTAAAGGAGAAAA
AAAAAATTCATTCCTGTAGGCGAGTGCCCGAACCCAAGATTTTTTTTTTTTTTTGAATTTTTTTCCTTTTTTTAAAAAAAGGGGGAG
AGAGGGGGGG

>R32_15_LB4_AD0_E09_071
CAANCGCATGNGTATTAGTTGTCTAAGCGTCAATTTGTTTACACCACAATATCTGCCCGTCGAGGATNGAACCAAGCCATTCGCTG
TGTTATTCTTCACGTTCGTCTGGTTCCGTCCTCGTGTCTAGCAAATTCCCTTCAACTGAGATGCAGCATGGCAAAGCGCGGTGATA
AAGACGTGAGCGTCCGGCGAAAGCGTTTTCGGGTATCGAGAGATGGAAGCTATGAGCCGCGTTTCATTTCATCGGCCATGACGCGG
TCGATCATTTCAGGATCGCACTGCGTGTCGATGAGGAGCTCGCGGATACCGCCGTCCCACGTCCTGATGTCGGCAAGTAGGCTTTG
AAGTTCTTCGATGCCATTCCCGGGCAAAACCCTTCGTGNCATCTTTCACTGGCATCGCTGAANGTAAAACCAGNTTCNCCCCCTCC
GGAAGANGTTGGTCCCGGGTTTTGNGGTTAACTTCTTTTCGAANAATTCNGAGGGTTTTCCCCNAANTNTTGGCCGGGGAAATTGG
GTTGAAACNGNANAAAAANGAAACCTNCGAGGGGNNGCCCCNTNNGGGGGGGGCCNCCCCNANNAANNNTTTTTNNTNNGGGGTTC
CCNGNTTTCCCNAAAAAANNCCCNAAATTTNTAAAAAAAAA

> R32_17-1_RB_AD6_E06
GACGGTAAATGTCGTTTCCGCCTTCAGTTTAACTATCAGTGTTTAAACTGATAGTTTAAACACTTGCGGATGACAAATTGACGCTT
GGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACTTAATAACACATTGCGGACGTTTTTAATGTACTGAATTAACGCCGAA
TTGCTCTAGCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAA
AGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGC
TAATTCGCTTCAAGACGTGCTCAAATCACTATTTCCACACCCCTATATTTCTATTGCACTCCCTTTTAACTGTTTTTTATTACAAA
AATGCCCTGGAAAATGCACTCCCTTTTTGTGTTTGTTTTTTTGTGAAACGATGTTGTCAGGTAATTTATTTGTCAGTCTACTATGG
TGGCCCATTATATTAATAGCAACTGTCGGTCCAATAGACGACGTCGATTTTCTGCATTTGTTTAACCACGTGGATTTTATGACATT
TTATGCTTCCGGGCTCGTATGG

> R32_17_LB_AD1_H03
AACTAATATAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACTACAATATATCCTGATCCA
AGCTTATTCCTAGGTTTTTCTGCTAGGCCCGTTTTTTTTCTAGTGTTCTACATCTTCCTTCCACGACACCACCATTTTTTCTCATC
CTAAATAGCTAGCTGTAAATTAAACCGTCGTCAAAACATGAAAAATGAAAAAAAAAACCTAGCCGATTTTGTTGTGCAATTAGTTGC
ACAAAACTACTTGATCTACAACTTTTTTTAAAAAATGGAATTATATCAGTTGAAGAAACTAAGTTCAATTTTACAAATTATTAACG
CAACTGTGTAGTTCTTATATATAAAAAAAAATAAACACCAGTTTAATATTTATATATTTGATCAATTAATCTACACCATTACAATT
TATTACAGTTAAGTCAGTTCTCTTAAGTTATTTACTCATTCCTTTGGTTTTGTACTTGGTTATACATTCCTGCACGAATGTTCGTA
ACCCATGACGTACGTTTTTGGAGAATAGTCATGTGTTATGCATGATCTTCATTCTAGCAAATAAACTATATATTCTTC

> R32_18-1_LB_AD0_H04
ACTGGGATATAGTAATTAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCAGCCTCATCCTTCTCCTC
CTCGCTTCTTCTCTCTTCTCCCTCCTCAGCGATGACAACGTCGATCTTGCTACCCTGGTTGCCTTCCTCTCTGCTGACCTCACCAA
TTCCGGCGGGCCCATGATGTCACCAGCCTTCACACTCTCACTGATGCCCTCCAATCTGGTGTAGGAAGGGGAAGAAGATGGAGGGC
GCAATTGGCATTGGGGACGTCATCGATTGGTTCTCTGGCAACCGGAGATGGGATCGCAGGAAGAGAGAGAGGAGGAGGATGGTGAC
CTTGGGGCAGAGGAGTCGAGGGAGGAGGATGGTGAATGTGGGGCGGAGTCATCGAGGAAGGGAGGGATAAGGTGGATGGCGACTGT
CTGACGAAGACGAGGGAGAAAACGGATCTGAGAAAACAGAAAATGGTTGGGTTGTATATTTATCCCGTAGCAGCAAAAACACCTGT
A

> R32_20-13_LB_AD1_H05
AATGGGGGTATAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACTTTCTTCGGTGGCGCGC
TGCTTCCTTCGATGGCTGGCGGGTGGCAGGCTGGTGGCTGGCAGTGGCAGGTGGCGGAGTAGCGGGTGGCGGCTGGTGGACGGGGA
GGAGTCGAGTACTCGAGCGGGTCATCGAGAGAAAGGGATCGGGAGGGAAGGGGATCGGCGGCCGATCGGGAGCGACGGTATTAGGG
AATCAGGAGAGGAGACTAGGGTTTTTGTTGGGCTGGGCCTCTAGCCCTTTATTATGTGACTGTGCGGCTAACACCCAAAGGCCCCC
CGCTCATATGTTCTCGCTCGTGAGCAGCTCGCGCGCCCGCTCGATCTCAAACGAGCTGACATGCCACCTTTTTTTTTTTTGGAAAC
CGAGTCGAATTTGTCACGATTTGAGGGGGCTGAAGAAACATGATT

> R32_21-1_LB_AD1_H06
AACGGGTATAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACACAATTTATCCCCAATATG
GCATCACTTGCTATTCATCCATACGCTGAAATCAGCCTTGTTTATTTCAGGAACATAACATTGAGCGCTGTATCAACTGGAAATCC
TTTTGCATAGAAAATGGCCTCAGCCTTCAGTTTACAAACTAGAAAAGCTAAAGGCTCGGTAAAATCCTTTTCCAAAAAGCAGCTCA
CCAGTCACCACCAAATTACCAATTGGTGTGTGATTCCGTCCTTTGGTTAGTGAAACTATGCATATGTACAAACAAGGAGGAACATC
GTGAAGGGGACTGTGCGGAGAAATTTTGCACACATATGGTACAAACCAAGGAGGAACATCATGAAGGGGACTATACGGATAGGATA
GGATGCGACTCGTTGGTTGTGGCAAATGACAAGCTCGGATGTGGGAAATGGTAGACTTTCTCACAAATCTTATCAACAACTGAAAA
TCAGTTTTAGTGTATTTTCTTTAGGCAAATGTGTACGGATACCATTGCTCCTTGGAGCTATAGTGCAGCATGTTAGCTGCGGAGTT
CACTGGAAAGATTTTTTTTTTTTTGGGGAAATAGTTAATTTAATTTCCTCGGGAAAGGGAACCACCCCCAAAAACCTATTTTTAGT
TTTTTTTCGGATAAAATCACATAAATA

> R32_22-24_LB_AD1_H08
ACTGGCATATAGTAATTAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACTCCTTCGGTGGCGCGCTG
CTTCCTTCGATGGCTGGCGGGTGGCAGGCTGGTGGCTGGCAGTGGCAGGTGGCGGAGTAGCGGGTGGCGGCTGGTGGACGGGGAGG
AGTCGAGTACTCGAGCGGGTCATCGAGAGAAAGGGATCGGGAGGGAAGGGGATCGGCGGCCGATCGGGAGCGACGGTATTAGGGAA
TCAGGAGAGGAGACTAGGGTTTTTGTTGGGCTGGGCCTCTAGCCCTTTATTATGTGACTGTGCGGCTAACACCCAAAGGCCCACAG
CTCATATGTTCTCGCTCGTGAGCAGCTCGCGAGCCAGCTCGATCTCTAACGAGCTGACATGCCAGCTCGGCTCGTTAGGAAACCGA
GCCGAGCTTGTCACGAGCCGAGCGAGCTAACGAGCCATGAGATTTCTGTCCACCCCTAGAGATGTCCCCGTCCCGTTCATCTCACT
CGACAAAA

> R32_32_LB3_AD9_C12
CAGGCCGGAGGGAGGGAGGGTTAATGAATGTGTGTATAAGAAGTTAAGGTAAATTAGTAAAATAAAAATAAGATGTGAAAGAGAGG
AGGGAATTAATATTAATCTCGACTTTGCACACATGACATAATGACATTGAGGCGTTATGAGTGAAAACTTAACAGTAACACTTGCG
AACTTTCGATGACATTGAAACAACCAATTTTCCAAATAACTATTGTGCCCCCCAAATTAAAAAACACACACGCAGCTAGCCAAGAT
GTCGTGCCAAATATATATCGACCTGTGCTTGAGTATTCCAGTAGTATTGTACAGCACGGCACCTGTAAACGGGCTCGCAGCGGGGT
GCGGGGTCCGGTTGGGTCAGAGCAGCAGAATCCCAAAATGGTCCCGGATTCTCTCATAACGAAACTGCCGCGAGTCGGGAACTTCC
CGGGTTGGTAAAAAGATTCCCCCCCCGTTTATGCAACAGTAGGAGCAAACCGGGTAGAATTAATACCCCTCTCCTCTCTCATAATT
AAGGACACCAATACGTTGACACGTTTTTTCCCGGGGCTGTTTTAAGAACCTACTAACCCCAATTCTCTACAATCCCCATGGAAGAA
GCCTGAATCATGCAAAGAGAATTTTACTATTGATATATATCATTGGAACTT