Differential pre-mRNA Splicing Alters the Transcript Diversity of *Helitrons*

Between the Maize Inbred Lines


* Department of Biological Sciences, Oakland University, Rochester, MI 48309-4401
§Department of Horticultural Sciences and Program in Plant Molecular and Cellular Biology, University of Florida, Gainesville, FL 32611

The sequences for *Hel1-331*, *Hel1-332*, *Hel1-333*, and their correlating transcripts are deposited under the GenBank accession numbers:

KF294013, KF294014, KF294015, KF294016, KF294017, KF294018, KF294019, KF294020, KF294021, KF294022, KF294023, KF294024, KF294025, KF294026, KF294027, KF294028, KF294029, KF294030, KF294031, KF294032, KF294033, KF294034, KF294035

© The Author(s) 2013. Published by the Genetics Society of America.
Running Title: Differential Splicing of Helitrons

Keywords: Helitron, alternative splicing, gene evolution, transposable element

1 Corresponding Author:
Shailesh Lal, Oakland University 3200 N. Squirrel Rd. Room 347 Dodge Hall of Engineering Rochester, MI 48309
Phone: (248) 370-2875
Email: lal@oakland.edu
ABSTRACT

The propensity to capture and mobilize gene fragments by the highly abundant Helitron family of transposable elements likely impacts the evolution of genes in Zea mays. These elements provide a substrate for natural selection by giving birth to chimeric transcripts by intertwining exons of disparate genes. They also capture flanking exons by read-through transcription. Here, we describe the expression of selected Helitrons in different maize inbred lines. We recently reported that these Helitrons produce multiple isoforms of transcripts in inbred B73 via alternative splicing. Despite sharing high degrees of sequence similarity, the splicing profile of Helitrons differed among various maize inbred lines. The comparison of Helitron sequences identified unique polymorphisms in inbred B73, which potentially give rise to the alternatively spliced sites utilized by transcript isoforms. Some alterations in splicing, however, do not have obvious explanations. These observations not only add another level to the creation of transcript diversity by Helitrons among inbred lines but also provide novel insights into the cis-acting elements governing splice site selection during pre-mRNA processing.
INTRODUCTION

*Helitrons* represent a novel family of highly abundant transposable elements in the genomes of species representing all eukaryotic kingdoms (Kapitonov and Jurka, 2001; Lal *et al*., 2003; Pritham and Feschotte, 2007; Yang and Bennetzen 2009; Lal *et al*., 2009a; Lal *et al*., 2009b). An autonomous *Helitron* is proposed to encode DNA helicase and a replication initiator protein. These are similar to bacterial transposons, which transpose via a rolling circle mechanism (Tavakoli *et al*., 2000; Kapitonov and Jurka, 2001). Typically, unlike other known transposable elements, *Helitrons* do not contain repeats of their terminal ends, and do not duplicate the host sequence upon insertion (Kapitonov and Jurka, 2001). Rather, they insert between host dinucleotide AT in the vast majority of cases, and bear conserved TC and CTRR nucleotides at their 5’ and 3’ terminal ends, respectively.

Although *Helitrons* are proposed to multiply via strand replacement and synthesis, similar to the copy and paste mechanism, recent PCR based detection of their somatic excisions in maize indicates these elements can also transpose via a cut and paste mechanism (Li and Dooner, 2009). Despite their importance, unequivocal genetic or biochemical evidence of their activity and movement to date remains elusive in any species. The presence of nearly identical copies of long *Helitrons* inserted in different regions of the genome points to their recent mobility in the maize genome. Our earlier report of two maize mutants caused by the recent insertion of *Helitrons* provided the first genetic evidence of their mobility in the present day maize genome (Lal *et al*., 2003;
Gupta et al., 2005). Recently, Helitrons were reported to be the most abundant DNA transposon in the maize genome. An improved version of the computational-based searches estimates the presence of more than 31,000 Helitrons. These constitute >6% of the total maize genome (Xiong et al., 2014). These Helitrons are prolific in capturing thousands of gene pieces and mobilizing them into different regions of the genome (Du et al., 2009; Yang and Bennetzen, 2009). Gene piece capture appears random, and the advantage of this to the host or the element remains unclear. The captured gene pieces display varying degrees of sequence similarity to their putative progenitors. It remains unclear whether this reflects the evolutionary time of their capture or natural selection.

Several contrasting mechanisms for gene capture by Helitrons have been proposed but each lacks supporting experimental evidence (Feschotte and Wessler, 2001; Brunner et al., 2005; Gupta et al., 2005; Kapitonov and Jurka, 2007; Lal et al., 2009a; Lal et al., 2009b). In addition to coding regions, Helitrons have mediated massive movement of a diverse collection of host sequences, including promoters, poly-A addition sites and binding sites for various regulatory proteins across the genome. Emerging reports indicate that in addition to coding regions, Helitrons have also captured and multiplied various host regulatory sequences that display functionality. These point to their possible important roles in physiological processes and adaptation of organisms to their environment (Han et al., 2013; Fu et al., 2013; Ellison and Bachtrog., 2013; Capriglione et al., 2014; Thomas et al., 2014).
Helitron-captured gene pieces represent dead remnants of their progenitors in the majority of cases. However, they are sometimes transcribed, giving birth to eclectic transcripts, which fuse coding regions of different genes. These may evolve into new genes with novel domains and functions following selection (Lal et al., 2003; Morgante et al., 2005; Brunner et al., 2005). The location of the promoters driving the expression of the captured genes remains elusive in the vast majority of cases (Morgante et al., 2005; Brunner et al., 2005; Jameson et al., 2008). We recently demonstrated that alternative splicing and read-through transcription dramatically augment the transcript diversity of Helitron-captured genes in maize (Barbaglia et al., 2012). These provide a potential substrate for natural selection. The molecular basis defining the aberrant alternative splicing of Helitron-captured genes is not apparent. Perhaps the close proximity of different gene sequences and mutations within the elements interfere with the proper recognition of their wild type splice sites. Mutations affecting splice site selection from a distance have been reported in both plants and animals (McNellis et al., 1994; Marillonnet and Wessler, 1997; Lal et al., 1999a).

Here, we performed expression analysis of selected Helitrons in different maize inbred lines. We previously reported alternatively spliced expression of these Helitrons in inbred B73 (Barbaglia et al., 2012). Our data indicate that these specific Helitrons inserted into these respective sites before the divergence of the inbreds examined here are expressed in both etiolated shoots and roots in other inbred lines. However, the pattern of alternatively spliced transcripts differed dramatically among the inbred lines. The
comparison of the Helitrons and their transcripts among the inbred lines identified key polymorphisms potentially affecting the differential usage of splice sites among inbred lines. These observations not only add another degree of complexity to the diversity of Helitron-captured genes among inbred lines for potential natural selection, but also provide novel insights into the mechanism of splice site selection during pre-mRNA processing in plants.
MATERIAL AND METHODS

Plant Material: The maize inbred lines B73, HP301, OH7B and Tzi8 were obtained from Maize Genetics Cooperative Stock Center, University of Illinois. Plants were grown in the greenhouse or in the field at the University of Florida/Institute of Food and Agriculture Sciences facility in Citra, FL.

Genomic and RT-PCR analysis: Genomic DNA was extracted from kernel tissue using the DNeasy Plant Mini Kit (Qiagen) following the protocol provided by the manufacturer. The PCR amplification of Helitrons Hel1-331, Hel1-332 and Hel1-333 in inbred lines HP301, OH7B and Tzi8 was achieved using the same primer pairs used for RT-PCR expression analysis of these Helitrons in maize inbred line B73 described previously (Barbaglia et al., 2012), except amplification of Helitron Hel1-333 was performed using two sets of overlapping PCR primers. The primer pairs H33-1F (5’-GCGTTCTGCGTTAGACAAT-3’) and H33-13R (5’-AGGGTGACCAAAGAGCAAGT-3’) spanning positions 51,979-52,018bp and 58,508-58,527bp, respectively, of HTGS clone (gi: 373839194) are complementary to exon 1 and exon 13 of predicted gene structure of the captured gene by Hel1-333. Similarly, primer pairs H33-13F (5’-ACCAGAGACGGAGGTCT-3’) and H33-14R (5’-TCATGCCCTTTACACTTGAT-3’) span positions 58,468-58,487bp and 60,207-60,226bp and complementary to exon 13 and exon 14, respectively. Primers designed on B73 inbred sequence failed to PCR amplify the border junction sequence of Hel1-331.
and *Hell*-332 in various inbred lines (data not presented). This may either be due to different insertion sites of *Hell*-331 and *Hell*-332 in different inbreds or due to the polymorphisms at the binding sites of the primers between the inbred lines. In contrast, the successful PCR amplification and sequencing of the products using the primers complementary to exon 14, which lies outside of *Hell*-333 strongly, indicates this *Helitron* is inserted at the same site in the inbred lines.

Total RNA was extracted from three-day-old, dark-grown etiolated roots and shoots of various inbred lines using Invitrogen’s TRIzol reagent. Resultant RNA was subjected to RT-PCR analysis using Superscript™ First Strand Synthesis kit (Invitrogen) using the same primers reported earlier for the expression analysis for *Hell*-331 and *Hell*-332 in inbred B73 (Barbaglia *et al*., 2012). However, a different set of primer pairs H33E1F (5’-GAGGCCACCGACACATATTC-3’) and H33E14R (5’-GCTTTTCCTGCTCACACCTTC-3’), spanning positions 51,865–51,855bp and 60,107–60,127bp of the HTGS clone (gi: 373839194) were used for RT-PCR analysis of *Helitron* *Hell*-333 in inbred lines HP301, OH7B and Tzi8. These primers hybridize to exon 1 and exon 14 of *Hell*-333 (Barbaglia *et al*., 2012). RT-PCR was performed using High Fidelity Platinum Taq Polymerase (Invitrogen) with at least three different RNA extracts from each inbred line to test the reproducibility of banding pattern of PCR products on agarose gels. The resultant PCR products from several different RT-PCR reactions were resolved on gel, purified, cloned into TOPO TA vectors (Invitrogen), and sequenced in both directions by the University of Florida Interdisciplinary Center for Biotechnology
Research DNA Sequencing Core Laboratory. Annotation and determination of splices sites were done by manual examination of the splice alignment of the *Helitrons* with their cognate RT-PCR amplified transcripts using the computer software GeneSeqer and SplicePredictor, respectively (Usuka and Brendel 2000a; Usuka *et al.* 2000b). The multiple sequence alignment to determine the polymorphic regions between the inbred lines was performed using the programs Multalin and ClustalW (http://multalin.toulouse.inra.fr/multalin/multalin.html) (Corpet, 1988; Larkin *et al.*, 2007).
RESULTS

Differential Splicing of Helitron-Transcribed Genes Among Various Maize Inbred Lines: We recently reported the expression analysis of several Helitrons in maize inbred line B73 (Barbaglia et al., 2012). The elements analyzed were initially selected from EST evidence of captured gene transcription. Our data indicated that these Helitrons produced multiple transcripts via differential selection of splice sites during pre-mRNA processing and read-through transcription. Since alternative splicing and Helitrons are both viewed as potentially important forces in the processes of evolution, we examined the expression of Helitrons Hel1-331, Hel1-332, and Hel1-333 in different maize inbred lines. Total RNA from etiolated root and shoot tissues from inbred lines B73, CML277, CML322, HP301, Ki11, Mo17, Ms71, OH43, OH7B and Tzi8 was subjected to RT-PCR analysis using the same primers for each Helitron reported previously in B73 (Barbaglia et al., 2012). The resultant amplified products were resolved on an agarose gel and stained with ethidium bromide. As displayed in figure 1, the profile of RT-PCR products for each Helitron in the other inbreds dramatically differed from B73. For example, several RT-PCR products from both roots (Figure 1, Panel A) and shoots (Figure 1, Panel B) were derived from Helitron Hel1-331 in inbred B73. In contrast, only a single product of approximately 1.0kb was amplified from roots in inbred lines CML277, CML322, HP301, OH7B and Tzi8 and shoots in inbred lines CML322, HP301, Ki11, Mo17, Ms71, OH43, OH7B and Tzi8 (Figure 1, Panel A and B).

As previously reported, multiple transcripts were derived from Hel1-332 in roots
and shoots of inbred B73 (Figure 1, Panel A and B) (Barbaglia et al., 2012). However, the profile of the amplified products differed considerably among the other inbred lines. For example, a single product of ~1.1kb was amplified in both roots and shoots of inbred line CML277. Similarly, a single product of ~0.9kb was detected in both roots and shoots of inbred line CML322. In contrast, HP301 produced three RT-PCR products of ~1.1kb, ~1.4kb and ~1.5kb in roots and a single product of ~1.3kb in shoots, whereas inbred Ki11 produced a single product of ~1.2kb in roots and two fragments of ~1.1kb and 1.3kb in shoots. No amplification was detected in roots of Mo17, although two fragments of 1.2kb and 1.8kb were amplified in shoots. Similar to Mo17, the inbred Ms71 produced two bands of ~1.4kb and ~1.8kb in shoots, whereas only a single band of ~1.2kb in roots. However, in contrast to Mo17, Ms71 amplified a single product of ~1.2kb in roots. Multiple products migrating between ~1.1kb and 1.4kb in length were amplified in both shoots and roots of inbred OH43. The inbred OH7B produced one fragment of ~1.3kb in roots and three fragments of ~1.1kb, 1.4kb and 1.8kb in shoots. A single product of ~1.2kb was amplified only in the roots of inbred Tzi8.

We reported previously that Helitron Hel1-333 had captured a flanking exon by read-through transcription and produced seven transcript isoforms by alternative splicing (Barbaglia et al., 2012). As displayed in both panels of figure 1, multiple products spanning between ~1.7kb and 2.2kb in length were amplified in both roots and shoots of inbred line B73. Similar to Hel1-332, the amplification profile of Hel1-333 diverged among different inbred lines. For instance, two major products of ~1.8kb and ~2.1kb
were amplified only in the roots of inbred line CML277. Inbred CML322 amplified a single product of ~1.9kb in roots and three distinct fragments of ~1.2kb, 1.3kb and 1.9kb in shoots. In contrast, HP301 produced a major product of ~2.0kb in roots and several weakly stained bands ranging between ~1.1kb and 1.7kb in length in shoots. However, the sequences of inbred HP301 shoot RT-PCR products displayed no similarity to Hel1-333 sequence, and were considered artifacts of PCR amplification (data not presented).

Amplification of inbred Ki11 RNA produced multiple fragments ranging from ~1.2kb to ~2.1kb in roots and from ~1.2kb to ~1.6kb in shoots. The inbred Mo17 contained two products of ~1.6kb and 1.9kb in shoots and only one product of ~1.9kb in roots. Multiple faintly stained and poorly resolved products ranging between ~1.0kb and ~2.0kb were amplified from Ms17 roots, and multiple products ranging from ~1.0kb to ~1.8kb in Ms17 shoot tissue. Poorly resolved, weak fragments ranging between ~1.2kb and 1.9kb were also amplified in both roots and shoots of inbred OH43. Amplicons from RNA of OH7B contained five fragments ranging from ~1.6kb to ~2.1kb in roots and two fragments of ~1.6kb and ~1.8kb in shoots, whereas inbred Tzi8 amplified a single product of ~1.9kb from roots, and two fragments of ~1.6kb and ~1.8kb from shoots.

We conclude that Helitrons Hel1-331, Hel1-332 and Hel1-333 are expressed in all inbreds reported here. The difference in the RT-PCR profile between inbred lines is caused by differences in splice site selection during pre-mRNA processing.

We selected inbred lines HP301, OH7B and Tzi8 to elucidate the molecular basis of differential splicing among maize inbred lines. These lines displayed markedly
divergent RT-PCR profiles of alternative splicing for the *Helitrons Hel1-331*, *Hel1-332* and *Hel1-333* as previously reported (Barbaglia et al., 2012).

**A Single Donor Splice Site Mutation Potentially Negates Alternative Splicing of *Hel1-331* transcripts in various maize inbred lines:** To elucidate the molecular basis of the observed differences in the splicing of *Hel1-331* between B73 and other maize inbred lines, we amplified the cognate *Hel1-331* from genomic DNA of lines HP301, OH7B and Tzi8 using primers designed for amplification in B73 (Barbaglia et al., 2012). The resultant product of ~2.9kb in length from each inbred line was cloned and sequenced in both directions. Supplemental figure 1 displays the multiple sequence alignment of *Hel1-331* between inbreds B73, HP301, OH7B and Tzi8. The inbred lines HP301, OH7B and Tzi8 shared ~99% sequence similarity between them, and each displayed ~95% similarity to inbred line B73. As illustrated, despite sharing high degrees of sequence similarity, distinct regions of polymorphisms span the entire length of the elements within these inbred lines. These included indels and small clusters of nucleotide changes in addition to single nucleotide changes. We also cloned and sequenced a single *Hel1-331* specific RT-PCR product amplified in both roots and shoots of HP301, OH7B and Tzi8. We then performed a splice alignment of each resultant transcript with their cognate inbred *Helitron*. Each line produced an identical gene structure with four exons and three introns (Figure 2, panel A). This predicted gene structure is identical to one, isoform IV, of the eight different isoforms encoded by *Hel1-331* in B73 (Barbaglia et al., 2012).
Noteworthy, the cognate region of the long intron 3 of this transcript encompasses the entirety of alternative splicing events and harbors the alternative donor and acceptor sites of the other seven transcript isoforms produced in B73 by *Hel1-331* (Barbaglia et al., 2012). Panel A of figure 2 schematically displays the position of these alternative donor and acceptor sites utilized to generate the other transcript isoforms of *Hel1-331* in B73.

Since sequences near splice site junctions play a critical role in splice site selection, we compared the flanking 15bp of the splice site junctions for *Hel1-331* transcript isoforms in B73 with the cognate region in inbred lines HP301, OH7B and Tzi8. Panel B of figure 2 displays the multiple sequence alignments of the splice site junctions of *Hel1-331* transcripts in B73 with their corresponding sequence in inbred lines HP301, OH7B and Tzi8.

A comparison of the donor and acceptor sites for intron 3 of this *Hel1-331* transcript from HP301, OH7B and Tzi8, with intron 3 of the B73 transcript isoform IV is displayed in sub-panels of figure 2B. There are three single nucleotide polymorphisms (SNP) between B73 and the other inbred lines. At one of these sites, a nucleotide C was substituted in three lines for a donor site G at -1 position in B73. Similarly, C replaced a donor site T at the +9 position in B73 and G replaced an acceptor site C at the -14 position in other inbred lines. The alignment of various B73-specific splice site junctions with other inbred lines are displayed in panels ii-vii of figure 2B. For example, the donor site at the B73 alternative splicing event ii is the same as the donor for alternative splicing event i. However, A replaced an acceptor site T at the +5 position located within exon 4.
in other inbred lines. The donor site for alternative splicing event iii displays no polymorphism between the inbred lines. However, the substitution of A to G at the -2 position alters the acceptor site dinucleotide from AG to GG in OH7B and Tzi8. Similarly, the donor site of alternative splicing event iv exhibits no changes, whereas the T in inbred B73 at position -15 of the acceptor site is replaced with A in other inbred lines. In addition, T in inbred HP301 replaces C, while the C remains unchanged in inbreds OH7B and Tzi8 at the -3 position of the acceptor site for inbred B73. The flanking sequences of splice sites used for alternative splicing event v were identical among the four inbred lines, except that C at +15 position of the acceptor site in B73 is replaced with T in other inbred lines. Similar to before, the splice sites of alternative splicing event vi are identical among four inbreds, except for an insertion of a G nucleotide at -12 position in other inbred lines and a T for an acceptor site C at +15 position in B73 in other three inbred lines. The flanking sequence of the donor site for alternatively spliced event vii remains identical among the different inbred lines, although it creates an acceptor site in exon 4, 95bp downstream to the 3’ splice site of intron 3. The comparison of the flanking exon 4 sequences of the other inbred lines to the acceptor site identified a single polymorphism. Here, a T replaced an acceptor site C at -9 position in B73 in inbred line OH7B.

**Various alternatively spliced transcript isoforms of Helitron Hel1-332 display inbred specific expression:** In the same manner, to decipher the molecular basis of RT-PCR
differences observed between these maize inbred lines, we cloned and sequenced a ~3.9kb fragment of *Helitron Hel1-332* PCR amplified from the genomic DNA isolated from inbred lines HP301, OH7B and Tzi8. Supplemental figure 1B displays the multiple alignment of the *Hell-332* sequence from inbred lines B73, HP301, OH7B and Tzi8. As shown, a high degree of similarity is shared between the *Hell-332* sequences of the four inbred lines even though regions of polymorphism, including indels, span the entire length of the element. The sequence of inbred line HP301 displayed highest similarity of >99% with B73, whereas OH7B shared 97% and Tzi8 had 94% similarity with B73. The alternative splicing produces six isoforms of *Hel1-332* transcript in inbred line B73 (Barbaglia et al., 2012). In addition, we sequenced the two products amplified from the inbred lines HP301 and OH7B and a single product from inbred Tzi8. We performed splice alignment of the resultant products with their corresponding *Helitron* in different inbred lines to determine the gene structure and position of the intron/exon junctions. The gene structure of a single product of *Hell-332* amplified from inbred Tzi8 is identical to transcript I of inbred B73 (Figure 2). Similarly, the two products amplified from HP301 displayed identical gene structure to transcripts I and II of inbred B73, while the two transcripts of OH7B were identical to transcripts II and V of B73. The transcripts III, IV and VI were unique to B73, and were absent in other inbred lines. The majority of splicing events that give birth to various transcript isoforms are concentrated in the 5’ end of the *Helitron*, spanning exon 1 and intron 1 of transcript I (Figure 2). The exception is the retention of an unspliced intron 5 of transcript I in transcript VI spanning the 3’ end.
of the *Helitron*. The alternative splicing events that utilize donor or acceptor sites different from transcript I are displayed in figure 2A and B. As shown, utilization of a non-canonical donor site within intron 1 and an acceptor site of exon 2 produces transcript II of B73. The flanking A at -10 position of the donor site in B73 was replaced with G in OH7B and Tzi8. Similarly, an alternative donor site within exon 1, an acceptor site of exon 2, and retention of an entire intron 1 give rise to B73 transcripts III and IV, respectively. As displayed in figure 2B, no polymorphism was detected flanking the splice site junction between the inbred lines of these splicing events. The use of several donor and acceptor sites creates two exons within intron 1 of B73 transcript V. There exists no polymorphism between the inbred lines flanking the splice junction of the first intron of this transcript. However, a flanking A at -2 position of the acceptor site of the second intron region in B73 was replaced by G in inbred Tzi8. In transcript vi, a G in inbred Tzi8 replaced T at -3 position of the acceptor site of the third intron region in B73. Similarly, T at +9 position of the acceptor site was replaced by C in inbred HP301. The retention of a complete intron 5 of B73 transcript isoform I produced transcript isoform VI. In transcript VI, the A at -15 position and T at -2 position in inbred B73 were both replaced by C in OH7B and Tzi8. The other splicing events common to all the six transcript isoforms of B73 did not display any junction site flanking polymorphism between the inbred lines.
Alternative selections of splice sites create transcript isoform diversity of *Hel1-333*

**between maize inbred lines:** The *Helitron Hel1-333* in maize inbred B73 encodes seven transcript isoforms via alternative splicing and read through transcription of a flanking exon spliced to the captured gene transcript (Barbaglia *et al.,* 2012). We amplified *Helitron Hel1-333* from inbred lines HP301, OH7B and Tzi8 using two pairs of overlapping primers. These primers are complementary to exon 1 and exon 13, and exon 13 and 14, respectively of inbred B73 *Hel1-333* sequence. The resultant products of 6.7kb and 1.9kb in length were cloned and sequenced. As shown in supplemental figure 1C, *Hel1-333* sequence from the three inbred lines share a high degree of similarity with inbred B73 *Hel1-333*. For example, HP301 and OH7B each display >97% sequence similarity to B73, while Tzi8 displays >99% similarity. However, changes like indels and single base changes interspersed throughout the length of the *Hel1-333* were detected between the inbred lines. RT-PCR using primers spanning exon 1 and the flanking exon 14 was performed on total RNA extracted from roots and shoots of different inbred lines. The resultant products were cloned, sequenced and spliced aligned to their cognate *Hel1-333* inbred line. Inbred HP301 produce a single *Hel1-333* specific transcript, whereas, OH7B and Tzi8, each produce three and two transcripts, respectively (Figure 4). Intriguingly, the splicing profile of each *Hel1-333* transcript amplified from inbreds HP301, OH7B and Tzi8 were distinct and did not display identity to any of the *Hel1-333* transcript isoforms of B73 (Barbaglia *et al.,* 2012). To elucidate the molecular basis of splicing differences, we compared the alternative splicing profile of *Hel1-333* transcript
isoforms of HP301, OH7B and Tzi8 with B73 *Hel1-333* isoform I as a reference transcript.

As displayed in figure 4, a single product amplified from HP301 revealed three regions of alternative splice site usage compared to inbred B73 *Hel1-333* transcript I. Use of an alternative acceptor site within intron 4 and donor site of exon 3 results in the skipping of exon 4, and increases the length of exon 5 by 44bp. The comparison of the splice site junctions of intron 3 and intron 4, and the alternative acceptor site region within intron 4, displayed no polymorphism between inbred B73 and HP301, regardless of the loss of exon 4 (Figure 4). However, we note an insertion of A at the -8 position of the acceptor site of downstream intron 5 in HP301, and an insertion of 9bp at position +12. In the same manner, the use of an alternative acceptor site within intron 7 and the donor site of exon 7 increases the length of exon 8 by 62bp. Despite this difference, the splice junctions of intron 7 exhibit no polymorphism between B73 and HP301. However, the acceptor splice site of downstream intron 8 displays a deletion of G at position +4. Also, an alternative use of an acceptor site within exon 13 in conjunction with the donor site of intron 12 increases the length of intron 12 by 61bp. The splice junctions of intron 12 and the alternative acceptor site within intron 12 are identical in B73 and HP301.

In the comparison between B73 and inbred OH7B, transcript I from OH7B displayed five regions of alternative splicing events. For example, use of an alternative acceptor site within exon 3 in combination with the donor site of intron 2 decreases the length of exon 3 by 5bp. The comparison of splice junction sequence of intron 2 and
flanking intron 1 and 3 between B73 and OH7B displays no polymorphism except for G at -15 position of the donor splice site of intron 1 in B73 is replaced by A in OH7B. Similar to the HP301 transcript, an alternative acceptor site within intron 4 results in the skipping of exon 4, and alters the length of exon 5 by 44bp. Except for a replacement of A at +5 position of the donor splice site of intron 4 in B73 with T in OH7B, the splice junctions of introns 3 and 4, and the alternative acceptor site remain identical between these two inbred lines. In addition, an alternative acceptor site within intron 7 increases the length of exon 8 by 44bp. We note the donor splice site of intron 7 has an insertion of A at +1 position in OH7B compared to B73. Also, an A in OH7B replaces a G at -1 position of the acceptor site of intron 8 in B73. In addition, utilization of an alternative donor and an alternative acceptor site creates an intron of 339bp within exon 10. The splice junctions of the upstream intron display no polymorphism; however, an A in OH7B replaces G at position +8 of the donor splice site of downstream intron 10 in B73, and C in OH7B replaces T at position -14 of the acceptor splice site of intron 10 in B73. An acceptor site within exon 13, in combination with donor site of intron 12, reduces the length of exon 13 by 61bp. The T, G and A at positions -3, +12 and +15 of the acceptor splice site of intron 12 in B73 is replaced by C, A and G, respectively, in OH7B. The splicing profile of OH7B transcript II is similar to OH7B transcript I with three exceptions. Unlike OH7B transcript I, the alternative acceptor splice sites within exon 3 and intron 7 are not evoked in transcript II. Also, the use of an alternative donor and an alternative acceptor site creates an intron of 316bp within exon 10. In OH7B transcript
III, the use of alternative non-canonical splice sites GC-TG within exon 3 and intron 13 leads to the skipping of exon 2 to exon 13. The splice junctions of this non-canonical intron display no polymorphic sequences between B73 and OH7B.

The Tzi8 transcript I displays three distinct regions of alternative splice site usage compared to B73 transcript I. Much like OH7B transcript I, an alternative acceptor site of intron 2 reduces the length of exon 3 by 5bp. The splice junctions of intron 2 remain identical in inbred lines B73 and Tzi8. Also, the employment of the donor site of intron 3, in conjunction with acceptor site of intron 4, results in skipping of exon 4. Except for a deletion of T at -4 position of the acceptor splice site of intron 3 in Tzi8, the splice junctions of flanking introns 2 and 4 remain identical between B73 and Tzi8.

Furthermore, similar to HP301 transcript I, an alternative usage of an acceptor site within exon 13, in conjunction with the donor site of intron 12, increased the length of intron 12 by 61bp. The splice junctions of intron 12, and the region flanking the alternative acceptor site with exon 13, display no polymorphism between B73 and Tzi8. Similarly, splicing pattern of Tzi8 transcript II is identical to HP301 transcript I, except usage of an alternative donor and an alternative acceptor site within exon 10 creates an intron of 339bp. We note the donor site of intron 9 displays an insertion of T at +9.
DISCUSSION

We previously reported that transcript diversity of Helitron-captured genes is dramatically augmented by alternative splicing (Barbaglia et al., 2012). These potentially provide a potent diversity pool for the evolution of new genes via natural selection. Here, we report the expression analysis of selected Helitrons, and demonstrate that differential usage of splice sites during pre-mRNA processing drastically alters the splicing profile and transcript isoform diversity between various maize inbred lines. We compared the sequences flanking the splice site junctions between the inbred lines to investigate the cis-acting elements potentially underlying splicing differences. The vast majority of splicing aberration reported in both plants and animals have occurred from mutations within the proposed consensus region around the splice site junction (Baralle and Baralle, 2005; Krawczak et al., 2007; Schuler, 2008). For example, a G to C transversion and a G to A transition at the +1 position of the donor splice site interferes with splicing and causes skipping of the downstream exon in both tomato and Arabidopsis, respectively (Jacobsen et al., 1996; Gorguet et al., 2009). Similarly, a base substitution at the -1 or -2 position of the donor splice site has also been reported to cause skipping of the downstream exon (Jack et al., 1992; McNellis et al., 1994; Sablowski and Meyerowitz, 1998; Lal et al., 1999b; Schuler, 2008), and the retention of unspliced intron or activation of a lariat-exon intermediate (Orozco et al., 1993; Liu and Filipowicz, 1996; Lal et al., 1999b; Choe et al., 2000; Yuan et al., 2009). Comparisons of splice junction sites of Helitron Hell-331 in our study revealed several base substitutions between B73 and other
inbred lines including a G to C transversion at the +1 position of the donor splice site of intron 3 in a single \textit{Hell-331} transcript. This substitution alters the donor site consensus from AG/GT to AC/GT in inbreds HP301, OH7B and Tzi8. Since the splice sites of all the seven B73 \textit{Hell-331} transcript isoforms reside in this intron 3 region, the most plausible explanation for the loss of alternative splicing in other inbred lines is this +1 position transversion substitution interfering with the recognition of their downstream splice sites. In this particular context, the various splice junction polymorphisms among the inbred lines detected at positions other than +1 of the donor splice site within intron 3, however, seem unlikely to play a role in splice site selection due to the following observation. We have shown that a single base substitution alters the acceptor site of B73 transcript isoform III from invariant terminal dinucleotide AG to GG in OH7B and Tzi8. Mutations that alter the acceptor site dinucleotides generally result in the complete abolishment of the splice site (Schuler, 2008), and as expected, the transcript homologous to B73 transcript isoform III does not occur in OH7B and Tzi8. However, the absence of the B73 transcript isoform III in inbred HP301, which does not bear the homologous splice site alteration, indicates this mutation does not impact splice site selection in other inbred lines and lends support to the +1 transversion substitution as the causative polymorphism.

Despite significant differences in the number of \textit{Hell-332} transcript isoforms between the inbred lines, each transcript isoform of \textit{Hell-332} detected in inbreds HP301, OH7B and Tzi8 was identical to one of the six transcript isoforms of \textit{Hell-332} reported
in B73 (Barbaglia et al., 2012). The inspection of the polymorphism flanking the splice sites between the inbred lines does not provide clear clues to their impact on the selection of splice sites. It is quite plausible that variance in alternative splicing may be caused by polymorphisms distantly located from the splice sites. The recognition of the splice sites appears to be dependent on gene context. Intragenic mutations influencing the selection of splice sites from a distance have been reported in both plants and animals (McNellis et al., 1994; Marillonnet and Wessler, 1997; Lal et al., 1999c). Intriguingly, each transcript isoform of Hel1-333 was unique in each inbred and displayed different combinations of splice site usage. Similar to Hel1-332, inspection of the splice junction failed to reveal significant polymorphisms potentially underlying the observed splicing differences. Whether read-through transcription and capture of the flanking exon have evoked employment of cryptic splice sites by various Hel1-333 transcript isoforms needs further investigation.

Unlike vertebrates, the mechanism of intron recognition and spliceosomal assembly is poorly understood in plants, mainly due to the lack of in vitro splicing assays. Despite sharing basic structural similarity, plant introns are not efficiently spliced in vertebrates and vice versa (Barta et al., 1986; van Santen and Spritz, 1987; Pautot et al., 1989; Barbazuk et al., 2008). Also, the presence of plant-specific splicing factors indicates intrinsic differences in splicing machinery between plants and animals (Barta et al., 2008; Rauch et al., 2014). Like vertebrates, the splicing differences of Helitrons among different inbred lines point to a complex array of signals that define the
mechanism of splice site recognition in plants. For example, growing evidence in vertebrates points to short five-to-ten nucleotide conserved elements located in both the introns and exons that affect splicing. Called SREs (Splicing Responsive Elements), they interact with various RNA binding proteins and may suppress or enhance the use of splice sites (Chen and Manley, 2009; Reddy et al., 2013). The mechanisms that regulate the use of splice sites remain uninvestigated in plants. The polymorphisms and mutations spanning the unidentified SREs may underlie the splicing differences of Helitrons between the inbred lines. In addition, the polymorphism may differentially impact the secondary structure of the Helitron-transcribed genes between the inbred lines.

Mutations that alter the secondary structure of RNA are well documented in vertebrates (Warf and Berglund, 2010; Wan et al., 2011), and have also been shown to affect splice site selection in plants (Goodall and Filipowicz, 1991), perhaps by disrupting the site for RNA binding proteins.

It is well documented that pre-mRNA processing in vertebrates does not occur independently of other processes. Rather it is linked to transcription where various splicing factors participate in splicing before termination of transcription. (Perales and Bentley, 2009; Moore and Proudfoot, 2009; Luco and Misteli, 2011). These studies demonstrate a strong link between the pattern of alternative splicing and elongation rate of RNAPII (Luco and Misteli, 2011; Kornblihtt et al, 2004; Luco et al., 2011). Since the elongation rate of RNA polymerase II (RNAPII) is controlled by chromatin structure, one explanation of the alternative splicing observed here is that Helitron-captured sequences...
of different genes have altered chromatin structure compared to their wild type progenitors. This may alter the elongation rate of RNAPII transcribing Helitron-captured genes. The change in rate of RNAPII may enhance/alter access to the splicing elements recognized by the spliceosome, or it may alter the recruitment profile of splicing factors. Altogether, we envisage this could generate spurious splicing events leading to the appearance of multiple spliced transcripts that we have shown to exist in Helitron-captured genes.

Whether the polymorphisms between the Helitrons among the various inbred lines occur randomly or under the selection to promote or suppress particular transcript isoforms remains to be determined. Clearly they may have important implications for the process of gene evolution. Perhaps the differentially spliced transcript isoforms of Helitrons provide adaptive advantages to plants grown in environmentally diverse regions of the world.

ACKNOWLEDGEMENTS

We thank Kathleen Lesich for reviewing the earlier version of the manuscript. This work was supported by the National Science Foundation (grant numbers 1221984 and 1412218) and by the Research Excellence Fund, Oakland University.
REFERENCES


Han, M. J., Y. H. Shen, M. S. Xu, H. Y. Liang, H. H. Zhang et al., 2013 Identification 
and evolution of the silkworm Helitrons and their contribution to transcripts. DNA 
Res. 20: 471-484.

of Arabidopsis thaliana encodes a MADS box and is expressed in petals and 

Jacobsen, S. E., K. A. Binkowski, and N. E. Olszewski, 1996 SPINDLY, a 
tetratricopeptide repeat protein involved in gibberellin signal transduction in 

Jameson, N., N. Georgelis, E. Fouladbash, S. Martens, L. C. Hannah et al., 2008 Helitron 
mediated amplification of cytochrome P450 monooxygenase gene in maize. Plant 


Kapitonov, V. V., and J. Jurka, 2007 Helitrons on a roll: eukaryotic rolling-circle 


Arabidopsis thaliana is a guanine-to-adenine transition at the 5'-splice junction of intron 3. Plant Physiol. 102: 227-232.


Figure Legends:

Figure 1. Expression analysis of Helitrons in different maize inbred lines. Panels A, B and C exhibit RT-PCR products extracted from roots (upper panels) and shoots (lower panels) of various inbred lines using primers specific for Helitrons, Hel1-331, Hel1-332 and Hel1-333, respectively. These Heltron-specific primers were based on the inbred B73 sequence described earlier (Barbaglia et al., 2012). The inbred lines are shown on top of the panel. Molecular weight markers in kilobases are displayed on the left of each panel.

Figure 2. Splicing patterns of Heltron Hel1-331 in different maize inbred lines. Panel A shows the gene structure of the Hel1-331 transcript isoforms detected in both roots and shoots of lines B73, HP301, OH7B and Tzi8. The boxes and lines show exons and introns, respectively. The magnified portion exhibits the positions of the donor and acceptor sites that generate the additional seven Hel1-331 transcript isoforms in inbred B73 (Barbaglia et al., 2012). The dotted lines connect the donor and acceptor sites. The orange line indicates the only transcript found in all four inbred lines. The black lines indicate the isoforms found only in B73. The Roman numerals i-vii mark these splice sites in the different transcript isoforms. Panel B presents the multiple sequence alignments of the splice junctions shown in panel A between the inbred lines. The exons and introns are displayed in upper and lower case letters, respectively. The
arrows connected by dotted lines indicate the donor and acceptor sites. The polymorphic nucleotides between the inbred lines are indicated by lettering and color.

**Figure 3. Alternatively spliced transcript isoforms of Helitron Hel1-332 display inbred-specific expression.** Panel A shows the gene structure of the six transcript isoforms I-VI of *Hel1-332* in inbred line B73 (Barbaglia et al., 2012). The transcript isoforms, specific to B73 and not detected in other inbred lines, are highlighted in green. The non-canonical splice sites are shown in blue, and asterisks mark the retained introns. Panel B exhibits the junction sequence alignment between the inbred lines of the splice sites marked by the lower case Roman numerals in panel A.

**Figure 4: Alternative selections of splice sites create transcript isoform diversity of Hel1-333 between maize inbred lines.** Panel A displays the schematic representation of the *Hel1-333* transcripts encoded by various inbred lines shown on the right. Panel B shows the splice junction alignment between the inbred lines of the alternative splicing events marked by Roman numerals in panel A.

**Supplemental Figure 1: Sequence Alignment of Helitron Hel1-331 between the lines of maize.** The sequence alignment exhibits the splice site junctions of various alternatively spliced transcript isoforms (i-vii) of *Hel1-331* in inbred B73 that are shown in figure 2. The arrows point to the donor and acceptor splice sites, and the blue and gray
shaded sequence flanking the splice sites represent exons and introns, respectively.
Asterisks beneath the highlighted regions indicate conserved flanking sequences between all the inbred lines. The names of the inbred lines are indicated on the left.

**Supplemental Figure 2: Alignment of Helitron *Hell*-332 between the maize inbred lines.** The alignment shows the splice junctions of alternative splicing events (i – viii) of *Hell*-332 shown in figure 3. The blue and gray shaded sequences flanking the splice sites represent exons and introns, respectively. Asterisks represent the conserved sequences beneath the highlighted regions, while the dashed lines fill the gaps in the alignment. The inbred lines are indicated on the left. The intron retention at splice sites iv and viii are marked by §.

**Supplemental Figure 3: Alignment of Helitron *Hell*-333 sequence between the maize inbred lines.** The sequence alignment displays the splice junctions i-xii of *Hell*-333 in B73 that are shown in figure 4. The blue and gray shaded sequences flanking the splice sites represent exons and introns, respectively. Conserved flanking sequences between all the inbred lines are indicated by asterisks beneath the highlighted regions, while the dashed lines fill the gaps in alignment. The inbred lines are indicated on the left.