Gene Capture by Helitron Transposons Reshuffles the Transcriptome of Maize

Allison M. Barbaglia,*1 Katarina M. Klusman,* John Higgins,*2 Janine R. Shaw,§ L. Curtis Hannah,§ and Shailesh K. Lal*

*Department of Biological Sciences, Oakland University, Rochester, MI 48309
§Department of Horticulture and Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL 32610–0245

1Present Address: Cell and Molecular Biology Program, Michigan State University, East Lansing, MI 48824-4320
2Present Address: Department of Engineering, Franklin W. Olin College of Engineering, Needham, MA 02492
Expression of *Helitron*-captured genes

*Helitrons*
Gene capture
Gene evolution
Alternative RNA splicing

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

*Helitrons* are a family of mobile elements that were discovered in 2001 and are now known to exist in the entire eukaryotic kingdom. *Helitrons*, particularly those of maize, exhibit an intriguing property of capturing gene fragments and placing them into the mobile element. *Helitron*-captured genes are sometimes transcribed giving birth to chimeric transcripts that intertwine coding regions of different captured genes. Here, we perused the B73 maize genome for high-quality, putative *Helitrons* that exhibit +/- polymorphisms and contain pieces of more than one captured gene. Selected *Helitrons* were monitored for expression via *in silico* EST analysis. Intriguingly, expression validation of selected elements by RT-PCR analysis revealed multiple transcripts not seen in the EST databases. The differing transcripts were generated by alternative selection of splice sites during pre-mRNA processing. Selection of splice sites was not random since different patterns of splicing were observed in the root and shoot tissues. In one case, an exon residing in close proximity but outside of the *Helitron* was found conjoined with *Helitron*-derived exons in the mature transcript. Hence, *Helitrons* have the ability to synthesize new genes not only by placing unrelated exons into common transcripts, but also by transcription read through and capture of nearby exons. Thus, *Helitrons* have a phenomenal ability to “display” new coding regions for possible selection in nature. A highly conservative, minimum estimate of the number of new transcripts expressed by *Helitrons* is approximately 11,000 or approximately 25% of the total number of genes in the maize genome.
INTRODUCTION

The Helitron family of transposable elements resides in the genome of species representing the entire eukaryotic kingdom (reviewed in Lal et al. 2009). While present in many genomes, the extent of their presence varies dramatically. In maize, the subject of these investigations, Helitrons compose ~2% of the total genome (Yang and Bennetzen 2009a; Du et al. 2009), Despite their massive abundance in several eukaryotic genomes, autonomous Helitron activity has not yet been reported in any species. The discovery of two maize mutants caused by recent insertions of Helitrons and the presence of nearly identical Helitrons at different locations in the maize genome point to their recent movement in maize (Lal et al. 2003; Gupta et al. 2005; Kapitinov and Jurka 2001; Lai et al. 2005). The detection of very recent somatic excisions of Helitrons in maize also indicates these elements are active in the present day maize genome (Li and Dooner 2009).

Helitrons are highly polymorphic in both length and sequence primarily due to different gene pieces captured by these elements (Du et al. 2009; Yang and Bennetzen 2009a; review by Feschotte and Pritham 2009). While several molecular mechanisms for gene capture have been proposed, (Feschotte and Wessler 2001; Brunner et al. 2005; Bennetzen 2005; Lal et al. 2009), definitive experimental evidence supporting a particular mechanism is still lacking. The capture of genes appears to be indiscriminate, and the biological relevance of capture to the element or the genome is not apparent. Captured genes exhibit varying degrees of sequence similarity to their wild type progenitors.

The massive diversity of Helitrons and their lack of terminal repeats as well as non-duplication of the insertion site sequences as associated with Class I and II transposable elements have made their detection computationally challenging. In maize, however, analysis of Helitrons associated
with +/- genetic polymorphisms identified a family of *Helitrons* containing conserved, short terminal ends. These conserved termini have been used to detect other family members (Gupta et al. 2005; Jameson et al. 2008). Recently, two computer-based programs, *Helitron*Finder and HelSearch, containing algorithms to recognize these terminal ends have been implemented to identify other *Helitrons* in the B73 genome (Du et al. 2008; Yang and Bennetzen 2009a; Yang and Bennetzen 2009b; Du et al. 2009). Both programs identified an overlapping set of ~2,000 putative, high-quality *Helitrons*. When these putative, high-quality elements identified using conserved terminal ends of the *Helitron* were used as a query in a blast search, an additional ~20,000 *Helitrons* or associated elements comprising ~2% of the total maize genome were identified (Yang and Bennetzen 2009a; Du et al. 2009). The vast majority of maize *Helitrons* have acquired gene fragments derived from up to 10 different genes embedded within a single element (Yang and Bennetzen 2009a; Du et al. 2009). These observations indicate that *Helitrons* have captured, multiplied, and moved thousands of gene fragments of the maize genome. How these events impact the evolution and expression of the maize genome is poorly understood. In comparison to *Helitrons* of other species, maize elements appear unique in their highly efficient ability to acquire gene fragments. This has significantly contributed to the diversity and lack of gene colinearity observed between different maize lines. This so called “+/− polymorphism” is primarily caused by presence and absence of gene-ferrying *Helitrons* between different maize inbred lines (Lai et al. 2005; Morgante et al. 2005).

The genes captured by *Helitrons* are sometimes transcribed, giving birth to eclectic transcripts intertwining coding regions of different genes. These potentially may evolve into new genes with novel domains and functions (Lal et al. 2003; Jameson et al. 2008; Brunner et al. 2005; Lal
and Hannah, 2005a; Lal and Hannah, 2005b; review Lal et al. 2009). Whether Helitrons have been a major driving force for gene evolution remains to be determined.

To analyze the transcriptional activity of Helitron-captured genes, we first identified highly reliable maize Helitrons in the sequenced B73 genome. These selected Helitrons had the following features: (1) contained terminal 5’ (5’-TCTMTAYTAMYHNW-3’) and 3’ (5’-YCGTNRYAAHGCACGKRYAHNNNCTAG-3’) sequences. These were derived from the multiple sequence alignment of the terminal ends of the Hell family of maize Helitrons (Dooner et al. 2007), (2) termini were in the correct orientation, (3) exhibited +/- polymorphisms in paralogs in B73 or in orthologs in other maize lines, (4) contained fragments of more than one captured gene, and (5) exhibited EST evidence of transcription. These Helitrons were further validated for their authenticity and the structure of their captured genes and transcripts by manual annotation. Resulting data indicate that Helitrons not only intertwine the coding regions of different captured genes but also generate multiple transcripts by alternative splicing and by read-through transcription that captures exons in genes near the Helitron. Hence, Helitrons are quite remarkable in generating diversity of coding regions which, upon selection, may lead to the evolution of new genes with novel domains and functions.

**MATERIAL AND METHODS**

**Plant Material:** The maize inbred lines described in this report were obtained from the Maize Genetics Cooperative Stock Center, University of Illinois. The plants were grown in the greenhouse or in the field at the University of Florida, IFAS facility at Citra Florida.
Identification of Helitrons and Expression Analysis of the Captured Genes: The conserved 5’ and 3’ terminal ends of the experimentally determined Hel1 family of Helitrons were isolated (Lal et al. 2008) and subjected to multiple sequence alignments. The strict consensus pattern of nucleotides displayed in Figure 1 was used as a template to search the entire database of Zea mays BAC sequences (B73 inbred) downloaded from the Plant Genome Database, Plant GDB (www.plantgdb.org/). A script was written in Python programming language using modules from the BioPython project to identify putative Helitrons. This program called HelRaizer, (secs.oakland.edu/helraizer) batch processes the input maize genome sequence and searches for sequences matching the terminal ends of the Helitrons. Correctly oriented 5’ and 3’ termini separated by 100bp to 25,000bp were identified and the intervening genomic sequence was labeled a putative Helitron. The identification of the Helitron captured gene fragments was performed using BLASTX search against the nr/protein National Center for Biotechnology Information (NCBI) database. Batch alignment was performed and alignments matching gene fragments of more than 50bp with at least 85% similarity were recorded as an instance of gene capture.

Evidence for movement of each putative Helitron from the screen above was sought by searching the B73 genome for a paralogous locus lacking the Helitron. This was determined by processing a 1,000bp sequence flanking each end of the element (minus the Helitron sequence) through the BLAST alignment against the Zea mays BAC sequence. In addition, the B73 genome was searched for sequences exhibiting significant internal sequence identity to the putative Helitron. Putative Helitrons from each of these two screens were monitored for expression. The putative duplicate elements that also shared sequence identity in their flanking BAC sequences were deemed redundant and were removed from the collection.
Expressed candidate Helitrons were identified by batch processing the putative Helitron sequences through the National Center of Biotechnology Information, NCBI (www.ncbi.nlm.nih.gov) BLAST (Basic Local Alignment Search Tool) analysis against the Expressed Sequence Tag (EST) database of Zea mays. Helitrons that had sequences aligning with the entire length of the EST sequences with at least 99% identity were assigned as candidates for expression of captured genes and were manually annotated and further pursued for experimental analysis. Figure 2 outlines the strategy used to discover Helitrons that display EST expression of captured host genes.

Annotation and structure analysis of captured gene pieces was done by manual examination of the splice alignment of the Helitrons with their cognate ESTs and their putative protein products using the computer software GeneSeqer (depc2.psi.iastate.edu/cgi-bin/gs.cgi) and SplicePredictor (depc2.psi.iastate.edu/cgi-bin/sp.cgi), respectively (Usuka and Brendel 2000; Usuka et al. 2000).

Genomic and RT-PCR Analysis: Genomic DNA extracted from kernel tissue of different maize inbred lines was performed using DNeasy Plant Mini Kit (Qiagen. Inc) according to the protocol provided by the manufacturer. Optimization of the PCR parameters for amplification in some cases was performed using a PCR optimization kit (Opti-Prime PCR, Stratagene, La Jolla, CA). PCR detection of +/- polymorphism of Hel1-331 (gi: 192757708; B73) between inbreds B73 and Mo17 was achieved using primers H31-1F (5’-CCGAATCTCACGTCGCTTAT-3’) and H31-1R (5’-AAGAGCCGGATAGCTTGACA-3’). These are complementary to positions 41,040bp – 41,060bp and 37,410bp – 37,430bp of the High Throughput Genomic Sequences
(HTGS) clone (GenBank accession AC220956) and span the 5’ and 3’ flanking sequence of the Hel1-331 insertion site, respectively. The RT-PCR analysis was performed on total RNA extracted from root and shoot tissues of maize inbreds B73 and Mo17 that were grown in the dark for three days using TRIZOL reagent (Invitrogen). The first strand was synthesized by oligo dT primers using SuperScript First Strand Synthesis System for RT-PCR (Invitrogen). Primer pairs, H31E1F (5’-AAGAGCCGGATAGCTTGACA-3’) and H31E7R (5’-ATATGCGCCAGGACAAGAAG-3’) were used for PCR amplification of Hel1-331. These primers are complementary to positions 44,230bp - 44,250bp and 41,656bp to 41,676bp of the HTGS clone (GenBank accession AC220956) and span exons 1 and 7, respectively, of the predicted gene structure by EST analysis. The RT-PCR analysis of Hel1-332a (gi: 209956049; B73) was performed on root and shoot B73 inbred RNA using primers H32E1F (5’-CGACAACCGATTTTCCAG-3’) and H32E6R (5’-GCCTACAACGATGGGCTAAT-3’), which are complementary to positions 145,787bp - 145,805bp and 149,498bp - 149,518bp of the HTGS clone (GenBank accession AC213839) and span exons 1 and 6 of the predicted gene structure by EST analysis. Similarly, primer pairs H33E1F (5’GAGGCCACCGACACATATTC-3’) and H33E14R (5’-GCTTTCCTGCTCACACCTTC-3’), complementary to exon 1 and exon 14 of EST predicted gene structure, were used for RT-PCR analysis of Hel1-333 (gi: 187358562; B73) on RNA isolated from B73 root and shoot tissue. These span positions 51,865bp – 51855bp and 60,107bp – 60,127bp of the HTGS clone (GenBank Accession AC205986). The RT-PCR of Hel1-334 (gi: 193211579; B73) used primers, H34E1F (5’-ATAGCGCTGGACACTTCCAC-3’) and H34E6R (5’-AGCGCCTGTATGGAGATGA-3’). These are complementary to exons 1 and 6 of the EST
predicted gene structure and span positions 116,802bp – 116,822bp and 120,472bp – 120,492bp of the HTGS clone (GenBank Accession AC211765), respectively.

The amplified PCR products were resolved on 1% agarose gels, excised, and purified using DNA agarose gel purification kit, QIAquick Gel Extraction Kit (Qiagen). The purified DNA was cloned and sequenced in both directions by either ABI Prism Dye Terminator sequencing protocol provided by Applied Biosystem, Foster City CA or done by the University of Florida Interdisciplinary Center for Biotechnology Research DNA Sequencing Core Laboratory.

RESULTS

Identification of Maize Helitrons Expressing Captured Genes:

We searched the B73 genome using the computer program, HelRaizer. This program predicts highly reliable Helitrons based on a strict consensus to the short, conserved terminal ends of the experimentally determined Hel1 family (Dooner et al., 2008). This program identified 2,376 putative Helitrons ranging from 168bp–5,024bp in length with an average and median length of 7,336bp and 6,129bp, respectively. These putative Helitrons compose 17.4Mb or ~0.73% of the total B73 genome. Sequences of 4,310 different gene fragments were detected within the predicted Helitron sequence, representing an average of 1.81 gene fragments per element. The preliminary analysis of the Helitrons discovered by HelRaizer displayed substantial overlap with the elements previously reported using other programs (Du et al., 2008; Du et al., 2009; Yang and Bennetzen, 2009a) (data not presented).

EST Evidence Indicates Expression of Two Genes Captured by Helitron, Hel1-331:
The alignment of *Hel1-331* (*gi: 192757708; B73*) with maize ESTs (gis: 71331232, 71324104, 71331231 and 78110425) predicted a gene structure of eight exons and seven introns embedded within the element (data not presented). The validation of *Hel1-331* was done by detecting +/- polymorphism for the insertion between inbreds B73 and Mo17. PCR amplification using primers flanking *Hel1-331* amplified a 344bp fragment from Mo17 DNA but not from B73 DNA (Figure 3A). The sequence of this amplified product indicated the presence of homologous regions differing by the presence of the *Hel1-331* (GenBank accession JN417509) insertion between nucleotides A and T in B73 (data not presented). From this observation and blastn analysis of the *Hel1-331* against the maize genome, we concluded that *Hel1-331* represents an authentic single copy Helitron insertion in inbred B73 but not in Mo17. The composite sequence of 2,127bp built from overlapping EST alignments produced an ORF of 307aa encoding the complete conserved domain of the nucleoside/nucleotide kinase superfamily of proteins and was identical to a hypothetical protein (*gi: 212721678*). The ORF also bore 98% sequence similarity to the carboxyl terminus of a maize heterogeneous nuclear ribonucleoprotein U-like protein 1, U1-hnRNP (*gi: 195655209*). The direct splice alignment of the U1-hnRNP protein with the *Hel1-331* element indicated a strong similarity to the first six exons of the EST predicted gene spanning 454aa residues of the 663aa carboxyl terminus of the U1-hnRNP protein, whereas, the last two exons revealed no similarity to known proteins in the database (Figure 3C). This observation indicates the transcript conjoins coding regions of two separate genes captured by this element.

*Hel1-331* Generates Multiple Transcripts Which are Differentially Spliced in Root and Shoot Tissue:
The RT-PCR analysis using primers complementary to exons 1 and 7 of the predicted gene amplified eight PCR products ranging ~700bp to ~2,300bp from root and shoot RNA from inbred B73 but not from Mo17 (Figure 3B). Fragments were cloned and sequenced. Figure 3C displays the schematic representation of the splice alignment of the resulting transcript sequences with *Helitron Heli1-331*. These transcripts are generated by differential selection of splice sites during pre-mRNA processing. For example, transcript I (GenBank accession JN638823) conforms to the gene structure predicted by EST evidence and contains seven exons ranging from 59bp – 888bp and six introns of 85bp – 322bp respectively. Transcript II (GenBank accession JN638824) retains intron 6, whereas transcript III (GenBank accession JN638825) retains both introns 3 and 6. Transcript IV (GenBank accession JN638826) is generated by utilization of a donor site of intron 3 and a cryptic acceptor site 95bp upstream to the acceptor site of intron 6, resulting in omission of exons 4, 5 and 6. Transcripts V (GenBank accession JN638827) and VI (GenBank accession JN638828) are generated by utilizing a cryptic donor and an acceptor site within exon 6, creating an additional intron of 544bp and 699bp, respectively within exon 6. Transcript VII (GenBank accession JN638829) is identical to transcript VI except it retains intron 6. Similarly, transcript VIII (GenBank accession JN638830) is identical to transcript VI but retains intron 3. Intriguingly, these alternatively spliced transcripts are differentially expressed in root and shoot tissues (Figure 3B). Inbred B73 roots exhibits three products of 1440bp, 1899bp and 2,221bp, corresponding to transcripts VIII, I and II, respectively. In contrast, B73 shoots produced six products of 938bp, 1,200bp, 1,355bp, 1,522bp, 1,899bp and 2,306bp. These correspond to transcripts IV, VI, V, VII, I and III, respectively. The predicted translation products encode proteins ranging from 189aa to 307aa residues. The multiple sequence alignment of these putative proteins as shown in figure 4
indicates that entire conserved domain of the nucleotide/nucleoside kinase superfamily remains intact in transcripts I, II, V and VI, whereas, transcripts III, IV and VIII lacks a minor portion of the amino terminal of the domain.

**Hell-332, a Member of a Helitron Gene Family, is Expressed:**

Comparison of a 1.4kb consensus sequence derived from the multiple sequence alignments of maize ESTs, gis: 78105127, 71450147, 18174728, 78105126, 8930323, 76909069 and 6021609 with the *Hell-332a* (GenBank accession JN638831) element revealed a gene structure containing six exons and five introns (data not presented). This 4,174bp element, *Hell-332a* (gi: 209956049; B73), spans positions 145,554bp-149,742bp contains portions of three different genes. The positions 170bp-645bp contained an ORF of 224 amino acid residues, which is annotated as an uncharacterized maize protein in GenBank (gi: 212275660). Similarly, a spliced alignment of a sorghum hypothetical protein (gi 242041151) bears sequence similarity to a five exon bearing gene structure spanning positions 1071bp-2751bp, whereas, positions 3779bp-3960bp displayed significant similarity to maize hypothetical protein (gi: 195657737) (Figure 5B). Four other members of the *Hell-332* family are: *Hell-332b* (gi: 166006896; B73) spanning position 132,003bp-136,174bp, *Hell-332c* (gi: 219689165; B73) spanning position 52,049bp-56,228bp, *Hell-332d* (gi: 221567066; B73) spanning position 27,404bp-31,607bp and *Hell-332e* (gi: 166852593; B73) spanning position 148,980bp-153,171bp. EST evidence for expression of other family members was not found.

**Alternative Splicing Produces at Least Six Populations of Hell-332 Captured Gene Transcripts:**
To validate the EST evidence of *Hell-332a* expression, we performed RT-PCR on total RNA from maize inbred B73 root and shoot tissues using primers complementary to exon 1 and exon 6 of the gene structure predicted by the spliced alignment of the maize ESTs with the *Hell-332a* element. The resulting RT-PCR products ranging from ~1,000bp to ~3,000bp from both root and shoot tissues were cloned and sequenced (Figure 5A). Of the eight cloned fragments, two lacked similarity to the *Hell-332a* and were discarded. The alignment of the resultant six sequences with *Hell-332a* (Figure 5B) indicates their origin by alternative splicing. For example, alignment of transcript I (GenBank accession JN638832) displayed six exons and five introns, which is identical to the gene structure predicted by the EST evidence. Transcript II (GenBank accession JN638833) utilizes an alternative donor and acceptor site inside intron 1 located 171bp downstream and 10bp upstream to the donor and acceptor site of intron 1, respectively. This creates a cryptic intron bearing non-canonical donor (TT) in combination with a non-canonical (AA) acceptor site within intron 1. Transcript III (GenBank accession JN638834) utilizes a cryptic donor site in exon 1, 233bp upstream to the donor site of intron 1 in combination with the acceptor site of intron 1. The entire sequence of intron 1 is retained in transcript IV (GenBank accession JN638835). The use of two alternative donor and acceptor sites creates two exons of 71bp and 344bp in length within intron 1 in transcript V (GenBank accession JN638836). Transcript VI (GenBank accession JN638837) is similar to transcript I except intron 5 is retained.

**Molecular and Expression Analysis of *Hell-333***:

The single copy *Hell-333* (*gi: 187358562; B73*) of 7,415bp in length, spanning position 51,355bp-58,769bp detected several paralogous loci precisely lacking the *Helitron* insertion
between dinucleotides A and T. A pairwise alignment of the sequence flanking the Hel1-333 (GenBank accession JN638842) insertion with one of the paralogous sequences, spanning position 179,556bp - 180,116bp of HTGS clone (GenBank Accession AC209160) is displayed in Figure 6A. Blastx analysis identified coding portions for three different proteins embedded within the Hel1-333 element. For example, position ~1,600bp – 1,800bp exhibited 85% similarity to a segment of a hypothetical protein (gi: 242043402) from sorghum. Similarly, position ~2,500bp – 6,900bp showed coding similarity to another hypothetical protein (gi: 242094646) from sorghum. SplicePredictor mediated a direct splice alignment of this protein with the Helitron sequence and detected 10 exons spanning the conserved peptidase domain within the element (data not presented).

The alignment of EST clones (gis: 224034606, 149102396, 76284017, 71768008, 76284017) all derived from a maize full-length cDNA library (Soderlund et al., 2009) with Hel1-333 and the flanking sequence, revealed a putative gene structure (PGS) consisting of 14 exons and 13 introns (Figure 6C, transcript I). Furthermore, the perfect alignment of these full-length ESTs within the 5’ boundary of the Helitron indicated they represent transcription initiation within the Helitron.

Intriguingly, the last exon of this EST is not contained within the Helitron, rather, this portion of the mRNA sequence was derived from a sequence just 3’ to the Helitron. This mRNA sequence shows perfect alignment with the flanking sequence of the 3’ boundary of the Helitron insertion creating an intron of 1,500bp in length and exhibits 93% similarity to a hypothetical protein (gi: 293335527) from maize. To validate the EST evidence, we performed RT-PCR on root and shoot RNA using primers complementary to exons 1 and 14 sequences, respectively. The amplified products (Figure 6B) were excised from the gel, cloned, and sequenced in both
directions. The alignment of the resulting sequences with \textit{Hel1-333} is shown in Figure 6C. These data indicate seven different transcript isoforms generated by alternative splicing. For example, transcript I (GenBank accession JN638843) aligns identical to the EST predicted gene structure. Transcript II (GenBank accession JN638844) revealed four regions of alternative splice site usage compared to the EST predicted gene structure. Use of an alternative acceptor splice site in intron 4 and donor site of exon 3, results in the complete skipping of exon 4. Similarly, usage of an alternative acceptor site inside intron 7 and donor site of exon 6 increases the length of exon 8 by 62bp. Also, alternative usage of both donor and acceptor sites creates an intron of 316bp internal to exon 10, and alternative acceptor site within exon 13 in conjugation with donor site of exon 12 decreases the length of exon 13 by 61bp. Transcript III (GenBank accession JN638845) utilizes a cryptic site downstream to the acceptor site of intron 2, thus decreasing the length of exon 3 by 5bp. Also, the usage of a donor site of exon 3 and the acceptor site of exon 5 results in skipping of exon 4, and a cryptic donor site internal to exon 10, in combination with the exon 11 acceptor site decreases the length of exon 10 by 502bp.

Transcript IV (GenBank accession JN638846) is generated by the combination of the splice sites described for transcripts I, II and III. For example, splicing from exons 1 – 7 follows the same pattern as transcript II, except for splicing of intron 2, which is similar to transcript III. Splicing of exons 7 – 10 follows the same pattern as transcript III, and splicing of exons 10 – 15 is similar to transcript I, except usage of alternative donor and acceptor site creates an exon of 50bp inside intron 12 and an alternative donor and acceptor site creates an intron of 315bp within exon 10.

Splicing of exons 1 – 7 of transcript V (GenBank accession JN638847) is similar to transcript II except usage of an alternative acceptor site within exon 7 increases the length of intron 6 by 62bp, and exons 7 – 14 is similar to transcript I, except for an alternative donor and acceptor site.
creating an intron of 439bp internal to exon 10. Similarly, splicing of exons 1 – 10 of transcript VI (GenBank accession JN638848) follows the same pattern as transcript V, except introns 8 and 9 remain unspliced, and splicing of exons 10 – 12 is similar to transcript II, except usage of an alternative donor site inside intron 11 increases the length of exon 11 by 8bp. Splicing of transcript VII (GenBank accession JN638849) follows a similar pattern to transcript II, except a usage of alternative acceptor site inside exon 7 and donor site of exon 6 decreases the length of exon 6 by 18bp, and the splicing of exon 9 is similar to exon 10 in transcript I. Intriguingly, all these alternatively spliced transcripts contained ORFs ranging from 84-105 amino acids residues in length that span the conserved peptidase domain (Figure 6C).

**Molecular and Expression Analysis of Hel 1-334:**

Another single copy Helitron, *Hel1-333* (GenBank accession JN638838) insertion of 4,492bp, spanning positions 116,272bp - 120,764bp in a maize HTGS clone (GenBank Accession AC211765) was discovered in chromosome 7. The authenticity of this element, *Hel1-334* (gi: 193211579; B73) was validated by the presence of a paralogous locus precisely lacking the *Helitron* insertion between the dinucleotides A and T (Figure 7A). The BLAST analysis of the element identified two regions spanning positions 315bp – 798bp and positions 1,751bp – 4,210bp with significant similarity to a hypothetical protein from sorghum (gi: 242080485) and an uncharacterized maize protein (gi: 226528348) (Figure 7C), respectively. The element lacked significant ORF to deduce biologically relevant function. The splice alignment of multiple overlapping maize ESTs produced a consensus structure of a gene containing six exons and five introns. The splice alignment of a representative EST (gi: 224031730) derived from a full-length cDNA clone and *Hel1-334* sequence is displayed in figure 7C (transcript I). The RT-PCR
analysis using primers complementary to exons 1 and 6 resulted in amplification products of approximately 400bp, 500bp, 1,000bp and 1,600bp in length using RNA template from both roots and shoots (Figure 7B). These fragments were excised, cloned, and sequenced. The alignment of the resulting sequences revealed three distinct alternatively spliced transcripts, each generated via alternative usage of the acceptor site of intron 1. For example, transcript I (GenBank accession JN638839) conforms to the gene structure predicted by EST evidence. In contrast, transcripts II (GenBank accession JN638840) and III (GenBank accession JN638841) utilized an alternative acceptor site 29bp downstream and 30bp upstream to the acceptor site of intron 1, respectively.

DISCUSSION

The abundance of Helitrons and their phenomenal ability to capture pieces of different genes and express them in chimeric transcripts strongly suggests that Helitrons are a major driving force in gene evolution. Analysis of the complete B73 genome sequence identified >20,000 Helitrons inserted primarily in gene rich regions (Du et al., 2009; Yang and Bennetzen, 2009a; Schnable et al., 2009; Feschotte and Pritham, 2009). These analyses also showed that maize Helitrons captured more than 20,000 gene fragments. Approximately 94% of these Helitrons contain exons derived from one to 10 different genes (Du et al., 2008, 2009; Yang and Bennetzen, 2009a). As we and subsequently others have reported, (Lal et al., 2003; Brunner et al., 2005; Lai et al., 2005) Helitrons shuffle exons and express these different captured genes in chimeric transcripts.
Here, we randomly selected four *Helitrons* and monitored their expression via RT-PCR analysis of RNA extracted from etiolated roots and shoots. In all cases, the *Helitron*-captured genes were transcribed into multiple transcripts generated via all known mechanisms of pre-mRNA splicing. These include exon skipping, intron retention, alternative selection of donor and acceptor splice sites, and non-canonical splice site selection. A total of 24 alternatively spliced transcripts expressed by these four elements were documented. Splicing is not random since splicing patterns observed in the root differed from those in the shoot. Also, it is interesting to note that the vast majority of the alternatively spliced transcripts reported here are not represented in the extant maize EST database. In this regard, we note that two maize genes, zmRSp31A and zmRSP31B, encode isoforms of arginine/serine (SR)-rich proteins via alternative splicing (Gupta et al., 2005). Similar to maize *Helitrons*, the majority of these transcript isoforms are not represented in the available maize EST collection (data not presented). Clearly the depth of maize ESTs is not sufficient to account for all the alternatively spliced events of the maize transcriptome.

While the retention of an unspliced intron in the mature transcripts of *Helitron* captured genes has been reported (Lal et al., 2003; Brunner et al., 2005), our data indicate that generation of multiple transcript isoforms via alternative splicing are quite widespread in expression of *Helitron*-captured genes. The impact of this process on maize genome evolution is dependent on the abundance and diversity of transcribed *Helitron*-captured genes. In this regard, we note that at least 9% of maize *Helitrons* exhibit extant EST evidence of expression (Yang and Bennetzen, 2009a). These studies suggest that of the ~20,000 high-quality *Helitrons*, ~1,800 elements are transcribed in at least one tissue (Yang and Bennetzen, 2009a). Here, we showed that only a small minority of the transcripts arising from *Helitron*-captured genes is currently present in
maize EST databases, hence, it is quite plausible that the vast majority of Helitron-transcribed sequences are alternatively spliced and the EST evidence of their expression may just represent the tip of the iceberg of their transcript diversity and abundance. Our data suggest Helitrons not only intertwine coding regions of different genes and transcribe them, but also augment the transcript repertoire by high levels of alternative splicing as well as capture of exon sequences from genes situated outside of the Helitron. Using the likely underestimate of expression from 1,800 Helitrons and our estimate of six transcripts arising from each Heliton-created gene, we estimate, at minimum, ~11,000 transcripts arise from Helitrons. It is highly implausible that these newly created sequences have not played a role in the evolution of maize genes and of maize.

We reported earlier the first case of incomplete splicing of exons from Helitron-captured genes. The splicing pattern appears to be determined contextually, and intragenic mutations acting from a distance to alter splice site selection occur in both plants and vertebrates (McNellis et al., 1994; Marillonnet and Wessler, 1997; Lal et al., 1999). It appears that reshuffling of exons originally residing in different genes changes the recognition of splice sites by spliceosomal machinery. How the new splice sites are recognized also appears to be tissue specific. For example, splice sites created by the insertion of the maize transposable element Dissociation (Ds) are recognized in the developing maize endosperm but not utilized in maize suspension cells (Lal and Hannah, 1999).

The aberration of transcript processing involving alternative splicing reported to date by transposable elements are caused by insertion of the element in either an exon or intron of the transcribed host gene (Wessler et al., 1987; Simon and Starlinger, 1987; Wessler, 1991; Ortiz and Strommer, 1990; Varagona et al., 1992; Chu et al., 1993; Giroux et al., 1994; Ruiz-Vazquez
and Silva, 1999). For example, insertion of Tgm-Express1, a member of CACTA family of transposable elements, in intron 2 of the Glycine max flavanone 3-hydroxylase (F3H) gene triggers alternative splicing of the mutant transcript. The resultant isoforms of the transcript display a unique combination of exons of five different gene fragments ferried by Tgm-Express1 spliced into F3H transcript (Zabala and Vodkin, 2007). Intriguingly, the analysis of the flanking sequence of all the Helitrons reported here indicates their insertion is not inside the transcribed regions of the host gene. In addition, the transcript appears to be initiated inside the element sequence.

The location of promoters driving transcription of captured genes inside the element has been proposed (Brunner et al., 2005; Morgante et al., 2005). For example, transcription of a maize cytochrome P450 monooxygenase captured by a Helitron seems to occur inside of the element (Jameson et al., 2008). In this regard, Helitrons are similar to pack-MULEs, where the initiation of transcription within the element is well documented (Jiang et al., 2004). In contrast, the promoter of the Sh2 gene drives the expression of the maize mutant sh2-7527 transcript containing the exons of different genes (Lal et al., 2003).

The perfect alignment of multiple ESTs derived from the full-length cDNA project within the element indicates that transcription is initiated inside the Helitron in all four cases reported here. The capture and splicing of a flanking exon located outside of the element with the transcript of captured genes initiated within the Helitron is intriguing, and to the best of our knowledge, has not been demonstrated with any other transposable element. This observation suggests that maize Helitrons, in addition to intertwining coding regions of different genes, dramatically increase their transcript diversity by alternative splicing as well as capture and splicing of flanking exon sequences. The abundance of Helitrons in genic-rich regions of the genome
suggests they are frequently flanked by exonic sequences that could potentially be spliced into the *Helitron*-transcribed sequences, thus, adding another dimension to further augment the diversity of transcripts created by these elements.

ACKNOWLEDGEMENT

This work was supported in part by National Science Foundation grant awards, 0514759, 0815104 and 1126267, USDA/NIFA grant, 2011-67003-30215 and by a Research Excellence Award, Oakland University.
FIGURE 1. Sequence Alignment of the Terminal Ends of Maize Helitrons: The left panel displays the names of the Helitrons: sh2-7527 (Lai et al., 2003), bal-Ref (Gallavotti et al., 2004), RplB73 (Gupta et al., 2005), ZeinBSSS53 (Song and Messing, 2003), P450B73 (Jameson et al., 2008), HelA-1 (Lai et al., 2005), HelA-2 (Lai et al., 2005), GHIJKLM9002 (Morgante et al., 2005), NOPQ9002 (Morgante et al., 2005), NOPQB73_14578 (Brunner et al., 2005), NOPQMo17_14594 (Brunner et al., 2005), NOPQB73_9002 (Brunner et al., 2005), Mo17NOPQ_14577 (Brunner et al., 2005), RST9002 (Brunner et al., 2005), U9002 (Brunner et al., 2005), HI9002 (Brunner et al., 2005), Hel-BSSS53+Zici (Xu and Messing, 2006), Hel-1 (Brunner et al., 2005), Hel-2 (Brunner et al., 2005), Hel-BSSS53-Zici (Xu and Messing, 2006), Hel-1 (Brunner et al., 2005). The middle and right panels exhibit the multiple sequence alignment of the conserved 5’ and 3’ termini of the Helitrons, respectively. The bottom panel shows the consensus sequence used for the database search for other Helitron family members.
FIGURE 2. Strategy Used to Discover Maize Helitrons and Analysis of their Captured Gene Expression: The upper panels exhibit the structure of non-autonomous maize Helitrons. The exons captured by non-autonomous Helitrons are represented by colored blocks. The terminal ends of the Helitrons are displayed by pattern filled boxes, and the loop near the 3’ terminus represents the palindrome sequence. The A and T nucleotides immediately flanking the insertion site of the Helitron are indicated.
FIGURE 3. Genomic and RT-PCR Analysis of Helitron Hel1-331: Panel A displays the PCR product amplified from genomic DNA extracted from different maize inbred lines using primers, H31-1F and H31-1R, flanking the 5’ and 3’ sequence of the Helitron insertion, respectively. Panel B shows the RT-PCR products amplified from root and shoot tissues of maize inbred lines B73 and Mo17 using primers, H31E1F and H31E7R. Panel C exhibits the splice alignment of the sequences of the RT-PCR products shown in panel B with the Helitron Hel1-331 sequence.
The exons of a captured hypothetical gene, gi: 212721678 and an uncharacterized gene are color coded in orange and yellow, respectively. In the alignment, boxes and lines denote exons and introns, respectively. Alternative donor and acceptor splice sites are joined by dashed lines and (*) marks the position of the retained introns. The size of the transcripts and the A and T nucleotides flanking the insertion site of the *Helitron* are indicated.
FIGURE 4. Protein Alignment of Alternatively Spliced Transcripts of *Hel1-331*: Alignment of the deduced protein sequences of *Helitron Hel1-331* transcripts are displayed in Figure 3C. The solid area marks the positions at which the same residue occurs in more than 60% of the sequences. The red line spans the conserved hnRNP-U1 domain.
FIGURE 5. Expression Analysis of Helitron Hel1-332a: Panel A displays the RT-PCR products resolved on a 1% agarose gel amplified from maize roots and shoots using primers E32E1F and E32E6R. Panel B exhibits the splice alignment of the Hel1-332a sequence with RT-PCR products shown in panel A. The boxes and lines denote exons and introns, respectively. Dashed lines join alternative donor and acceptor sites and (*) denotes a retained intron. The sizes of the RT-PCR products are indicated on the right. The captured gene fragments of proteins, gi: 212275660, gi: 242041151 and gi: 195657737 are displayed in green, blue and violet, respectively.
FIGURE 6. Molecular and Sequence Analysis of Helitron Hell-333: Panel A exhibits a pairwise sequence alignment of HTG sequence flanking the Hell-333 insertion (upper sequence) with the paralogous locus. An arrow marks the putative insertion site of the Helitron. Panel B displays the RT-PCR products from maize roots and shoots amplified using primers H33E1F and H33E14R. The splice alignment of the RT-PCR products in panel A with the Hell-333 sequence is shown in panel C. The boundaries of the Helitron and the predicted length of the RT-PCR products are indicated. The (*) marks the retained intron and alternative donor and acceptor sites are joined by dashed lines. The gene fragments of proteins, gi: 242043402, 242094646, and 29333527 are color coded in red, fuchsia and pink, respectively. The fuchsia shaded regions of the exons of the alternatively spliced transcripts represent the ORFs spanning the conserved peptidase domain.
FIGURE 7. Genomic and RT-PCR Analysis of *Helitron* Hell-334: Panel A displays a pairwise sequence alignment of the flanking HTGS (upper sequence) without the *Helitron* insertion and the sequence of the paralogous locus. The putative insertion site of the *Helitron* is marked by an arrow. Panel B shows the RT-PCR products amplified from root and shoot tissues using primers H34E1F and H34E6R. Panel C provides a schematic representation of the exon and intron junction of the alternatively spliced products in panel B. Exons of the captured genes, gi: 242080485 and gi: 226528348 are color coded in lime green and aqua, respectively. The dashed lines join alternative donor and acceptor sites. The predicted sizes of the transcripts are indicated.
REFERENCES

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

Brunner, S., G. Pea, et al., 2005  Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize. Plant J. 43: 799-810.

Chu, J. L., J. Drappa, et al., 1993  The defect in Fas mRNA expression in MRL/lpr mice is associated with insertion of the retrotransposon, ETn. J Exp Med. 178: 723-730.

Dooner, H. K. and L. He, 2008  Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. Plant Cell. 20: 249-258.

Dooner, H. K., S. K. Lal, et al., 2007  Suggested guidelines for naming Helitrons in maize. Maize Genet. Coop. News Lett. 81: 24-25.

Du, C., J. Caronna, et al., 2008  Computational prediction and molecular confirmation of Helitron transposons in the maize genome. BMC Genomics. 9: 51.

Du, C., N. Fefelova, et al., 2009  The polychromatic Helitron landscape of the maize genome. Proc Natl Acad Sci USA. 106: 19916-19921.

Feschotte, C. and E. J. Pritham, 2009  A cornucopia of Helitrons shapes the maize genome. Proc Natl Acad Sci USA. 106: 19747-19748.

Feschotte, C. and S. R. Wessler, 2001  Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. Proc Natl Acad Sci USA. 98: 8923-8924.

Giroux, M. J., M. Clancy, et al., 1994  De novo synthesis of an intron by the maize transposable element Dissociation. Proc Natl Acad Sci USA. 91: 12150-12154.

Gupta, S., B. B. Wang, et al., 2005  Two novel arginine/serine (SR) proteins in maize are differentially spliced and utilize non-canonical splice sites. Biochim Biophys Acta. 1728: 105-114.

Gupta, S., A. Gallavotti, et al., 2005  A novel class of Helitron-related transposable elements in maize contain portions of multiple pseudogenes. Plant Mol Biol. 57: 115-127.

Jameson, N., N. Georgelis, et al., 2008  Helitron mediated amplification of cytochrome P450 monooxygenase gene in maize. Plant Mol Biol. 67: 295-304.

Jiang, N., Z. Bao, et al., 2004 "Pack-MULE transposable elements mediate gene evolution in plants." Nature 431: 569-573.
Kapitonov, V. V. and J. Jurka, 2001  Rolling-circle transposons in eukaryotes.  Proc Natl Acad Sci USA. 98: 8714-8719.

Lai, J., Y. Li, et al., 2005  Gene movement by Helitron transposons contributes to the haplotype variability of maize.  Proc Natl Acad Sci USA. 102: 9068-9073.

Lal, S., J. H. Choi, et al., 1999  The AG dinucleotide terminating introns is important but not always required for pre-mRNA splicing in the maize endosperm.  Plant Physiol. 120: 65-72.

Lal, S. K., M. J. Giroux, et al., 2003  The maize genome contains a helitron insertion.  Plant Cell. 15: 381-391.

Lal, S. K. and L. C. Hannah, 1999  Maize transposable element Ds is differentially spliced from primary transcripts in endosperm and suspension cells.  Biochem Biophys Res Commun. 261: 798-801.

Lal, S. K. and L. C. Hannah, 2005a Helitrons contribute to the lack of gene colinearity observed in modern maize inbreds.  Proc Natl Acad Sci USA. 102: 9993-9994.

Lal, S. K. and L. C. Hannah, 2005b Plant genomes: massive changes of the maize genome are caused by Helitrons.  Heredity. 95: 421-422.

Lal S.K., M. Oetjens, and L.C. Hannah, 2009 Helitrons: Enigmatic abductors and mobilizers of host genome sequences. Plant Science 176:181-186.

Lal S.K., N. Georgelis and L.C. Hannah, 2008 Helitrons: their impact on maize genome evolution and diversity. The Maize Handbook: Domestication, Genetics, and Genome. Eds. J Bennetzen and S Hake.

Li, Y. and H. K. Dooner, 2009 Excision of Helitron transposons in maize.  Genetics. 182: 399-402.

Marillonnet, S. and S. R. Wessler, 1997 Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing.  Plant Cell. 9: 967-978.

McNellis, T. W., A. G. von Arnim, et al., 1994 Genetic and molecular analysis of an allelic series of cop1 mutants suggests functional roles for the multiple protein domains.  Plant Cell. 6: 487-500.

Morgante, M., S. Brunner, et al., 2005 Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize.  Nat Genet. 37: 997-1002.
Ortiz, D. F. and J. N. Strommer, 1990  The Mu1 maize transposable element induces tissue-specific aberrant splicing and polyadenylation in two Adh1 mutants.  Mol Cell Biol. 10: 2090-2095.

Ruiz-Vazquez, P. and F. J. Silva, 1999  Aberrant splicing of the Drosophila melanogaster phenylalanine hydroxylase pre-mRNA caused by the insertion of a B104/roo transposable element in the Henna locus.  Insect Biochem Mol Biol. 29: 311-318.

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

Simon, R. and P. Starlinger, 1987  Transposable element Ds2 of Zea mays influences polyadenylation and splice site selection.  Mol Gen Genet. 209: 198-199.

Soderlund, C., A. Descour, et al., 2009  Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs.  PLoS Genet. 5: e1000740.

Song, R. and J. Messing, 2006  Gene expression of a family in maize based on noncollinear haplotypes.  Proc Natl Acad Sci USA. 100: 9055-9060.

Usuka, J. and V. Brendel, 2000  Gene structure prediction by spliced alignment of genomic DNA with protein sequences: increased accuracy by differential splice site scoring.  J Mol Biol 297: 1075-1085.

Usuka, J., W. Zhu, et al., 2000  Optimal spliced alignment of homologous cDNA to a genomic DNA template.  Bioinformatics. 16: 203-211.

Varagona, M. J., M. Purugganan, et al., 1992  Alternative splicing induced by insertion of retrotransposons into the maize waxy gene.  Plant Cell. 4: 811-820.

Wang, Q. and H. K. Dooner, 2006  Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus.  Proc Natl Acad Sci USA. 103: 17644-17649.

Wessler, S. R., 1991  The maize transposable Ds1 element is alternatively spliced from exon sequences.  Mol Cell Biol. 11: 6192-6196.

Wessler, S. R., G. Baran, et al., 1987  The maize transposable element Ds is spliced from RNA.  Science. 237: 916-918.

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

Yang, L. and J. L. Bennetzen, 2009a  Distribution, diversity, evolution, and survival of Helitrons in the maize genome.  Proc Natl Acad Sci USA. 106: 19922-19927.
Yang, L. and J. L. Bennetzen, 2009b  Structure-based discovery and description of plant and animal *Helitrons*. Proc Natl Acad Sci USA. **106**: 12832-12837.

Zabala, G. and L. Vodkin, 2007  Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in Glycine max. BMC Plant Biol. **7**: 38.