Helitron mediated amplification of cytochrome P450 monooxygenase gene in maize

Natalie Jameson · Nikolaos Georgelis · Eric Fouladbash · Sara Martens ·
L. Curtis Hannah · Shailesh Lal

Received: 23 July 2007 / Accepted: 22 February 2008
© Springer Science+Business Media B.V. 2008

Abstract The mass movement of gene sequences by *Helitrons* has significantly contributed to the lack of gene collinearity reported between different maize inbred lines. However, *Helitron* captured-genes reported to date represent truncated versions of their progenitor genes. In this report, we provide evidence that maize CYP72A27-Zm gene represents a cytochrome P450 monooxygenase (P450) gene recently captured by a *Helitron* and transposed into an *Opie-2* retroposon. The four exons of the CYP72A27 gene contained within the element contain a putative open reading frame (ORF) for 428 amino acid residues. We provide evidence that *Helitron* captured CYP72A27-Zm is transcribed. To identify the progenitor gene and the evolutionary time of capture, we searched the plant genome database and discovered other closely related CYP72A27-Zm genes in maize and grasses. Our analysis indicates that CYP72A27-Zm represents an almost complete copy of maize CYP72A26-Zm gene captured by a *Helitron* about 3.1 million years ago (mya). The *Helitron*-captured gene then duplicated twice, approximately 1.5–1.6 mya giving rise to CYP72A36-Zm and CYP72A37-Zm. These data provide evidence that *Helitrons* can capture and mobilize intact genes that are transcribed and potentially encode biologically relevant proteins.

Natalie Jameson and Nikolaos Georgelis contributed equally to this manuscript.

N. Jameson · E. Fouladbash · S. Martens · S. Lal (✉)
Department of Biological Sciences, Oakland University,
Rochester, MI 48309-4401, USA
e-mail: lal@oakland.edu

N. Georgelis · L. C. Hannah
Department of Horticultural Sciences and Program in Plant
Molecular and Cellular Biology, University of Florida,
Gainesville, FL 32611, USA

Keywords Gene capture · Genome evolution ·
Helitrons · Transposable elements

Introduction

Transposable elements play a major role in the evolution of eukaryotic genomes (Feschotte et al. 2003; Curcio and Derbyshire 2003; Bennetzen 2005). Despite their wide abundance, until recently virtually all known transposable elements were grouped into two classes based on their mechanistic mode of transposition. Class I elements transpose via RNA intermediates catalyzed by reverse transcriptase enzyme encoded by an autonomous member of the family. Elements of Class II transpose via DNA intermediates catalyzed by transposases encoded by autonomous elements.

Despite the different mode of transposition, both classes of transposable elements contain either direct or inverted repeat sequences on their terminal ends and, with few notable exceptions, virtually all make a family-specific duplication of host site sequence upon insertion. Recently a novel family of eukaryotic transposable elements, termed *Helitrons* was discovered. These bear no apparent similarity to known Class I and II groups of elements. These were initially discovered by computational analysis of the complete genome sequence of Arabidopsis and *C. elegans* (Kapitonov and Jurka 2001). *Helitrons* have also been reported from other plants, invertebrates, fungi and fish (Kapitonov and Jurka 2003; Lal et al. 2003; Poulter et al. 2003). *Helitrons* also reside in mammalian genomes. Recently two families of these elements have been reported to constitute ~3% of the brown bat *Myotis lucifugus* genome (Pritham and Feschotte 2007).

Despite their wide abundance, these enigmatic elements remain undetected because, unlike generic transposable elements, *Helitrons* do not contain conserved sequences easily detected by computer based database searches. Furthermore, unlike other transposable elements, *Helitrons* do not contain repeat sequences at their termini, nor do they make duplications of host sequence upon insertion. Although an autonomously active *Helitron* has not been isolated, it is postulated that they encode DNA helicase and replicase/ligase protein similar to bacterial rolling circle transposon. Presumably they duplicate by a mechanism involving replication and strand replacement (Kapitonov and Jurka 2001; Feschotte and Wessler 2001; Poulter et al. 2003). Recently, we described two maize mutants that were caused by insertion of *Helitron* type transposable elements. This provided first tangible evidence that an active *Helitron* resides in the modern day maize genome (Lal et al. 2003; Gupta et al. 2005).

A fascinating feature of some transposable elements is their ability to capture portions of genes, increase their copy number and position them at multiple sites within the genome. For example, the LINE1 and non-LTR family of retrotransposons created thousands of pseudogenes inserted into different regions of the human genome (Zhang and Gerstein 2004). The portions of cellular genes embedded within both Class I and II groups of transposable elements often reported in plants are generally regarded as infrequent incidents of gene capture (Talbert and Chandler 1988; Jin and Bennetzen 1994; Elrouby and Bureau 2001; Kawasaki and Nitasaka 2004). However, a recent report of the capture of more than 1000 cellular genes by *Mu* family of transposable elements suggests that this process may have played a far more important role in plant genome evolution (Jiang et al. 2004). We earlier described several *Helitron* type transposable elements in maize which, unlike *Helitrons* reported from other species, completely lacked sequence similar to helicase or replicase/ligase (Lal et al. 2003; Gupta et al. 2005). Characterized maize *Helitrons* are laden with portions of multiple genes apparently captured by these elements during their movement across the genome. Recently, *Helitron*-mediated movement of cellular genes was shown to massively change the maize genome and cause a lack of gene colinearity between different maize inbred lines (Lai et al. 2005; Morgante et al. 2005; Lal and Hannah 2005a, b; Bennetzen 2005).

While significant progress has been made in characterizing *Helitrons*, the nature of genes ensnared by them remains poorly understood. *Helitron*-captured maize gene segments described so far are small gene pieces that contain in-frame stop codons. *Helitron* capture of an almost intact coding region of a maize gene for cytidine deaminase was reported recently (Xu and Messing 2006). The captured gene segment lacks exon 1, a large portion of intron 1 and

presumably its promoter. While the truncated coding regions and intervening intron sequences have apparently been conserved, efforts to detect expression of this captured gene segment have been unsuccessful. Here, we report a recent *Helitron* capture and amplification of a maize cytochrome P450 monooxygenase (P450) gene into three different regions of the genome. One of these amplified P450 genes within the *Helitron* contains an uncorrupted open reading frame of 428 amino acids. This encodes the highly conserved heme binding domain of P450 enzymes. We provide EST evidence that this captured gene is transcribed. These data provide further evidence that *Helitrons* can mobilize genes leading to genetic heterozygosity among inbred lines.

Materials and methods

Plant material

The maize inbred lines were obtained from Maize Genetic Cooperation Stock Center, University of Illinois. Plants were grown in the greenhouse or in the fields of the University of Florida, Gainesville.

Database searches

The 50 bp sequence of each terminus of *Helitron* insertion in maize mutants, *sh2-7527* (gi: 1321098) and *bal-ref* (gi: 49659739) were used as queries to search maize Non-Redundant (nr) and High Throughput Genomic Sequence (HTGS) deposited in the Genbank as of July 4th, 2007 (Altschul et al. 1990) (<http://www.ncbi.nih.gov/>). The clones displaying both termini of either mutant insertion were further perused for *Helitrons* (Lal et al., 2008). The intervening sequence between the termini was considered a *Helitron* insertion.

The annotation of the captured genes embedded within a *Helitron* insertion was determined by spliced alignment of the available ESTs, cDNAs and protein sequences using the computer software GeneSeqer and SplicePredictor available on their World Wide Web sites (<http://bioinformatics.iastate.edu/cgi-bin/gs.cgi>) and (<http://bioinformatics.iastate.edu/cgi-bin/sp.cgi>), respectively (Usuka and Brendel 2000; Usuka et al. 2000).

The CYP72A27-*Zm* cDNA (Genbank accession number: AF465266) was used as query to search Genbank, Maize Assembly Genomic Islands (MAGI), Sorghum Assembly Genomic Islands (SAMI) and Center for Integrative Genomics (CIG) for homologous genes in grasses. The nucleotide sequences of these genes were subjected to multiple sequence alignment using Bioedit software (BLOSUM62). The gene tree was constructed using Maximum likelihood (ML) analysis performed by the

GARLI (Genetic Algorithm for Rapid Likelihood Inference) software (Zwickl 2006). These analyses used the default General Time Reversible (GTR) model of sequence evolution (Yang 1994) by assuming some sites are invariant and the rates at the remaining sites are gamma distributed (Gu et al. 1995). Bootstrap values were also calculated by GARLI using 100 replicates.

The pairwise nucleotide distance (K) between introns of CYP72A26-*Zm* and CYP72A27-*Zm*, CYP72A36-*Zm*, CYP72A37-*Zm* was calculated by Kimura's (1980) two-parameter method using MEGA 3.1 program (Kumar et al. 2004). The time of divergence between CYP72A26-*Zm* and the group of CYP72A27-*Zm*, CYP72A36-*Zm*, CYP72A37-*Zm* was calculated by the formula: $T = (K_{\text{CYP72A26-Zm} - \text{CYP72A27-Zm}} + K_{\text{CYP72A26-Zm} - \text{CYP72A36-Zm}} + K_{\text{CYP72A26-Zm} - \text{CYP72A37-Zm}}) / 3 \times 2 \times (\text{average synonymous nucleotide substitution per site per year})$. The time of divergence between CYP72A27-*Zm* and the group of CYP72A36-*Zm*, CYP72A37-*Zm* was calculated by the formula: $T = (K_{\text{CYP72A27-Zm} - \text{CYP72A36-Zm}} + K_{\text{CYP72A27-Zm} - \text{CYP72A37-Zm}}) / 2 \times 2 \times (\text{average synonymous nucleotide substitution per site per year})$. The average synonymous nucleotide substitution per site per year for maize is around 6.5×10^{-9} (Gaut et al. 1996).

Genomic and RT-PCR analysis

Genomic DNA from various inbreds was extracted using Plant DNAzol solution (Invitrogen Inc.) following manufacturer's instruction. Optimum conditions for PCR amplification in some cases were achieved using a PCR optimization kit (Opti-Prime PCR, Stratagene, La Jolla, CA). Detection of the genomic sequences with the *Helitron* insertion was achieved by separate amplification of the 5' and 3' regions of the *Helitron* and their flanking sequences. Amplification of the 5' end of the *Helitron* was achieved using primers (5P450gF1, 5'-TTTGAAGTCCACGACAC-CAG-3') and (5P450gR1, 5'-TCGTACAAATTGGCAGA TGG-3'). These are complementary to positions, 113,393–113,413 and 113,790–113,809, respectively of the maize BAC clone (gi: 42491468) and span the 5' junction of the *Helitron*. Similarly, the 3' end of the *Helitron* was amplified using primer sets (3P450gF1, 5'-TGGCTTGAACA TGGTAGGTG-3') and (3P450gR1, 5'-TCGACCTCAACAAAGGAGGT-3'). These correspond to positions 119,426–119,446 and 119,901–119,921, respectively and span the 3' *Helitron* junction. Amplification of the maize locus containing the *pol* gene of *Opie-2* retroposon without the *Helitron* insertion was done using primers, (P450F1, 5'-GCTCACTACGCAGTCTGCAA-3') and (P450R1, 5'-TGTGGCCGTTAAGAGGATTC-3'). These primers span positions, 113,056–113,057 and 119,750–119,770 and complementary to the sequence flanking the 5' and 3'

regions of the *Helitron*, respectively. RT-PCR was performed on total RNA extracted from maize B73 roots and shoots grown in the dark for three days using TRIZOL reagent (Invitrogen). Oligo dT-primed first strand was synthesized using SuperScript First Strand Synthesis System for RT-PCR (Invitrogen). Primer pairs, P450F1 and P450R1 were used to analyze the expression of *pol* gene of *Opie-2* retroposon without the *Helitron* insertion. Primer pairs, 5P450gF1 and 5P450gR1 which flank the 5' end of the *Helitron* were used to detect possible genomic DNA contamination in RNA preparation. No product was produced with these primers. Resultant PCR products were resolved on 1% agarose gels and purified using DNA agarose gel purification kit, GENECLEAN II (Bio101, Inc). The purified products were cloned using TOPO TA Cloning Kit (Invitrogen) and were sequenced in both directions using the ABI Prism Dye Terminator sequencing protocol provided by Applied Biosystems, Foster City, CA.

Results

Maize CYP72A27-*Zm* represents a cytochrome P450 monooxygenase gene recently captured by a *Helitron*

Despite extreme sequence divergence and size polymorphisms, most characterized maize *Helitrons* bear short stretches of conserved terminal ends. These termini have been used to search for other family members in the maize genome database (Gupta et al. 2005; Lal et al. 2008). Using similar methodologies, we discovered a 50bp region spanning positions 1161–1211 of a genomic clone (gi: 21842135) that was nearly identical to the 5' termini of the *Helitron* insertions in *sh2-7527* and *bal-ref* (Lal et al. 2003; Gupta et al. 2005). The sequence of the genomic clone (gi:21842135) from position 1607 to 3273 shows 100% identity to cDNA (gi: AF465266) sequence and contains three introns. This cDNA is annotated as a maize partial gene for cytochrome P450 monooxygenase (P450) and is termed here CYP72A27-*Zm*. Immediately upstream of the *Helitron* insertion site, starting at position 1160, is an open reading frame of 154 amino acids for a protein with greater than 95% similarity to the carboxy terminus of the *pol* protein of the maize retroposon *Opie-2* (gi: 7489803). These observations suggested that the *Helitron* containing CYP72A27-*Zm* had inserted into the *pol* gene of an *Opie-2* retroposon. To further investigate this possibility, we searched the maize genome database and discovered an HTGS clone (gi: 48491468) that exhibited 100% sequence identity to the maize CYP72A27-*Zm*. This BAC sequence not only contained the 5' *Helitron* terminus described above but also contained approximately 30 bp sequence with strong similarity to the 3' terminal sequence of the

maize *Helitron* insertions described previously (Lal et al. 2003; Gupta et al. 2005). This region of similarity, from positions 119,648 to 119,677, was located 6,188 bp downstream of the putative 5' *Helitron* insertions site. It is immediately flanked by a downstream nucleotide T. These observations suggest that maize BAC gi: 42491468 contains a *Helitron* insertion spanning position 113,489 to 119,677. The *Helitron* location in the BAC clone and the sequence alignment of the terminal ends of the *Helitron* are shown in Fig. 1. The genomic PCR amplification and sequencing of both the terminal ends of the *Helitron* and its flanking sequence further confirmed the presence of this *Helitron* in maize B73 inbred (Fig. 2). We named this *Helitron* *Hel1-16* (gi:42491468;B73) following proposed nomenclature guidelines for maize *Helitrons* (Dooner et al. 2006).

The presence of the *Helitron* as an insertion sequence was also confirmed by the discovery of four overlapping maize ESTs, gis: 22490638, 19057786, 38568866 and 37361139. They display 99% sequence similarity to the sequence immediately flanking the *Helitron*, but lack the intervening 6,188 bp. The blastx analysis of these ESTs displayed more than 96% similarity to the *pol* protein of maize retroposon *Opie-2*, gi: 18542174. To further test the EST evidence for *Helitron* insertion, we performed PCR analysis of the maize inbred B73 genomic DNA and RT-PCR analysis of the mRNA isolated from B73 roots and shoots using primers flanking the 5' and 3' ends of the insertion site of the *Helitron*. These resulted in the amplification of a single 527 bp fragment from both genomic and mRNA samples (Fig. 3a). The sequence of the PCR product displayed more than 97% similarity to the BAC flanking sequences. The position of the *Helitron* insertion site between the dinucleotide A and T is marked by an

arrow in the pairwise sequence alignment between the PCR product and the BAC sequences (Fig. 3b). These results indicate that B73 DNA contains *pol* genes of *Opie-2* retroposon both lacking and bearing the *Helitron* insertion.

The *Helitron* insertion contains portions of several other genes not related to CYP72A27-Zm

The maize *Helitrons* described to date are embedded with portions of multiple genes apparently captured during their transposition through the genome (Lal et al. 2003; Gupta et al. 2005; Lai et al. 2005; Morgante et al. 2005). To further ascertain the presence of other cellular genes, we performed alignments of the available 451,122 maize ESTs with the complete *Helitron* *Hel1-16* (gi:42491468;B73) insertion using the computer software GeneSequer (Usuka et al. 2000). GeneSequer predicts the position of the exon/intron boundaries of a given genomic segment based on the strength of the assigned splice sites. The results of the alignment are schematically displayed in Fig. 4. A total of 104 matching maize ESTs were detected. The EST alignment also produced several PGS (Predicted Gene Structures). Four non-overlapping PGS were deemed reliable from manual inspection of the EST alignments. These are labeled PGS 1-4 in Fig. 4. PGS 1, which corresponds to maize CYP72A27-Zm (sequences 113,940 to 115,609 has four exons and three introns. Similarly PGS 2, (position 116,033 to 116,619) contains an open reading frame for 130 amino acids that encodes a highly conserved motor domain found in kinesin proteins. This participates in spindle assembly and chromosome segregation during cell division (Vale and Fletterick 1997). PGS 3 (117644 to 118424 exhibits weak similarity to oxidase/dehydrogenase (gi: 56784387) from tobacco whereas PGS 4 (118769

Fig. 1 Structure of maize *Helitron*, *Hel1-16* (gi:42491468;B73) Panel A displays the structure of *Hel1-16* (gi:42491468;B73) insertion in maize HTG clone, gi: 42491468. The element and flanking sequences are displayed by black and red lines, respectively. Exons of genes corresponding to CYP72A27-Zm, kinesin-like spindle protein, proline oxidase and leucine zipper factor contained within the

element are displayed by black, green, brown and yellow boxes, respectively. The arrows mark the position of the primers used during PCR analysis. Panel B displays the alignment of 5' and 3' terminal ends of *Hel1-16* (gi:42491468;B73) with *Hel-BSS53-z1C1* (Xu and Messing 2006) and *Helitron* insertion in maize mutants, *sh2-7527* and *bal1-refl*

Fig. 2 PCR amplification of the *Helitron* insertion: (a) The lower and upper case letters denotes *Hel1-16* (*gi:42491468;B73*) and flanking BAC sequences, respectively. Arrows mark the position of the PCR primers. (b) Ethidium bromide stained 1% agarose gel of the PCR products amplified from inbred B73 DNA using primers displayed in

panel (a). The molecular weight standards are displayed on the right. Primers 5P450gF1 and 5P450gR1 amplify the 5' end, and primers 3P450gF1 and 3P450gR1 amplify the 3' end of the *Helitron* insertion. Primers P450F1 and P450R1 amplify the maize locus containing the *pol* gene of *Opie-2* retroposon without the *Helitron* insertion

to 119462) has strong similarity to a leucine zipper factor like protein from rice. These observations suggest that in addition to CYP72A27-*Zm* gene, portions of at least three other genes are embedded within the *Helitron* insertion. Furthermore, the direct splice alignment of the matching proteins with the *Helitron* sequence indicated that the PGS 2-4 represent portions of corresponding gene sequences that no longer encode functional proteins.

Maize CYP72A2-*Zm* gene likely represents an intact portion of CYP72A26-*Zm* gene captured by a *Helitron*

In order to estimate the time of capture and movement of CYP72A27-*Zm* gene by a *Helitron*, we gleaned other CYP72A27-*Zm* related genes in maize and other grasses. The

names and database source of these candidate genes are listed in Table 1. To determine the progenitor of CYP72A27 gene, we performed a phylogenetic analysis of CYP72A27-*Zm* related genes in grasses. The gene tree in Fig. 5 shows that maize CYP72A36-*Zm* and CYP72A37-*Zm* followed by CYP72A26-*Zm* are the closest relatives of CYP72A27-*Zm*. Furthermore, both maize CYP72A36 and CYP72A37 genes are adjacent to the 5' conserved terminal end of *Helitrons* and share more than 98% sequence identity with CYP72A27-*Zm*. The total divergence in the sequence immediately flanking their *Helitron* 5' ends indicates that CYP72A27-*Zm*, CYP72A36-*Zm* and CYP72A37-*Zm* represent genes captured by a *Helitron* and inserted into different regions of the maize genome. Contrary to CYP72A27-*Zm* that contains a putative open reading frame for 428 amino acids, the open reading

Fig. 3 PCR analysis of *Helitron-16* (*gi:42491468;B73*) insertion site in *Opie-2* retroposon. **(a)** Ethidium bromide stained PCR products resolved on 1% agarose gel amplified from mRNA isolated from maize roots and shoots, and from inbred B73 genomic DNA. The size of the PCR products is indicated on the right. **(b)** Pair-wise alignment

of the BAC sequences flanking the *Helitron* (upper sequence) and the genomic PCR product amplified with primers, P450F1 and P450R1 from genomic DNA isolated from maize inbred B73 line (lower sequence). The insertion site of the *Helitron* is marked by an arrow

Fig. 4 Alignment of the maize ESTs with *Helitron-16* (*gi:42491468;B73*) in the alignment, introns and exons are displayed by lines and boxes, respectively. Maize ESTs, predicted gene structures and open reading

frames of more than 64 amino acids are displayed by red, green and orange bars, respectively

Table 1 CYP72A27-*Zm* related genes in different plant species

The table lists the annotation, accession number and databases source. Abbreviations and web address for databases in the table are Genbank [US National Center for Biology Information (NCBI); website: <http://www.ncbi.nlm.nih.gov>], MAGI [Maize Assembly Genomic Islands (Iowa State University); website: <http://magi.plantgenomics.iastate.edu/blast/blast.html>], SAMI [Sorghum Assembly Genomic Islands (Iowa State University); website: http://magi.plantgenomics.iastate.edu/blast/blast_s.html] and CIG [Center for Integrative Genomics (Cornell University); website: <http://www.phytozome.net/sorghum>]

Plant species	Tentative gene annotation	Accession number	Database source
Maize	CYP72A26- <i>Zm</i>	AY071866	GenBank
Maize	CYP72A27- <i>Zm</i>	AF465266	GenBank
Maize	CYP72A36- <i>Zm</i>	Magi_5103	MAGI
Maize	CYP72A37- <i>Zm</i>	Magi_5106	MAGI
Maize	CYP72A38- <i>Zm</i>	Magi_38892	MAGI
Maize	CYP72A39- <i>Zm</i>	Magi_72160	MAGI
Maize	CYP72A40- <i>Zm</i>	Magi_38906	MAGI
Sorghum	CYP72A41- <i>Sb</i>	Samiv2_56028	SAMI
Sorghum	CYP72A42- <i>Sb</i>	Super_2: 7388652..7386670	JGI
Rice	CYP72A43- <i>Os</i>	NM_001050172	GenBank
Rice	CYP72A44- <i>Os</i>	NM_001050166	GenBank
Rice	CYP72A45- <i>Os</i>	NM_001050168	GenBank
Rice	CYP72A46- <i>Os</i>	AB038597	GenBank
Rice	CYP72A47- <i>Os</i>	NM_001050033	GenBank
Wheat	CYP72A48- <i>Ta</i>	AF123605	GenBank
Barley	CYP72A49- <i>Hv</i>	DQ838490	GenBank

frames of CYP72A36-*Zm* and CYP72A37-*Zm* are interrupted by in frame stop codons negating coding capacity of a biologically active protein. Maize CYP72A26-*Zm* represents a full length version of P450 and shares 95% sequence identity to CYP72A27-*Zm*. Both genes contain three introns whose positions are evolutionary conserved (Fig. 6a). The pair-wise alignment of the putative translation products of the two genes are displayed in Fig. 6b. The alignment shows the presence of additional 99 amino acid residues at the amino terminus of CYP72A26-*Zm* gene. Based on the high degree of similarity between the two proteins, it seems CYP72A27-*Zm* lacks 99 amino terminus residue encoded by the cognate exon 1 of CYP72A26-*Zm* gene.

The phylogenetic analysis indicates that CYP72A26-*Zm* is the closest non-*Helitron* relative of CYP72A27-*Zm*.

Hence, the most parsimonious explanation is the progenitor gene of CYP72A27-*Zm* is CYP72A26-*Zm*. The *Helitron*-captured CYP72A27-*Zm* gene then duplicated twice, giving rise to CYP72A36-*Zm* and CYP72A37-*Zm* (Fig. 7). By calculating the rate of nucleotide substitution in introns as described in Material and Methods, we estimated the time of divergence between maize CYP72A26-*Zm* and the group of CYP72A27-*Zm*, CYP72A36-*Zm* and CYP72A37-*Zm*. It appears that an intact portion of CYP72A26-*Zm* gene was copied and inserted into a different region of the genome around 3 million years ago (mya) giving birth to CYP72A27-*Zm* gene. The *Helitron*-captured CYP72A27-*Zm* gene then duplicated twice to give rise to CYP72A36-*Zm* and CYP72A37-*Zm* around 1.5–1.6 mya. Whether the two duplication events occurred simultaneously or

Fig. 5 Phylogenetic analysis of CYP72A27-*Zm* related genes in grasses. Bootstrap values >50% are shown, and the scale bar indicates the number of nucleotide substitutions per site

Fig. 6 Comparison of CYP72A26-Zm with CYP72A27-Zm. (a) Gene structures of CYP72A26-Zm and CYP72A27-Zm. The exons and introns are displayed by boxes and lines, respectively. The 5' terminal end of the *Helitron* is marked by an arrow. The region (marked by red

line) between the 5' terminus of *Helitron* and exon 1 of CYP72A27-Zm is totally divergent from CYP72A26-Zm. The length (bp) of the exons and introns are indicated. (b) The pair-wise alignment of the putative translation products of CYP72A26-Zm and CYP72A27-Zm

sequentially cannot be determined because of the close time proximity between the duplication events giving rise to CYP72A36-Zm and CYP72A37-Zm genes.

Maize ESTs specific to CYP72A27-Zm, CYP72A36-Zm and CYP72A37-Zm are present in extant databases

Portions of genes captured by *Helitrons* are sometimes transcribed (Lal et al. 2003; Lai et al. 2005; Morgante et al. 2005). The resulting transcripts are usually chimeric

Fig. 7 *Helitron* mediated amplification of a P450 gene in maize: The progenitor maize P450 gene duplicated into CYP72A26-Zm and CYP72A38-Zm. An almost complete copy the maize CYP72A26-Zm gene was captured by a *Helitron* about 3 mya. The *Helitron*-captured gene then duplicated twice, approximately 1.5–1.6 mya. The designations CYP72A27A-Zm, CYP72A27B-Zm and CYP72A27C-Zm were arbitrarily assigned

consisting of cojoined sequences of several genes. It has been postulated that these chimeric transcripts evolve to encode novel proteins. To investigate these possibilities, we pursued maize ESTs displaying 100% sequence identity to each of the P450 genes contained within each *Helitron*. Maize ESTs gis: 76015821, 71761546, 78085771 and 78086969 derived from the full-length cDNA library prepared from mRNA of various maize tissues and developmental stages displayed 100% identity CYP72A27-Zm gene. Similarly, EST gi:148956719 shows 100% identity to CYP72A36-Zm and EST gi: 71771445 and 91055659 show 100% identity to CYP72A37-Zm. From these observations we conclude that *Helitron* captured maize P450 genes CYP72A27-Zm, CYP72A36-Zm and CYP72A37-Zm are transcribed.

Discussion

Origin of the CYP72A27-Zm, CYP72A36-Zm and CYP72A37-Zm genes

Here we describe a new *Helitron* which contains a gene, CYP72A27-Zm. This gene may encode cytochrome P450 monooxygenase. The most parsimonious explanation is that CYP72A27 gene was captured by the *Helitron*. However, we cannot formally exclude the remote possibility that this P450 gene moved into the *Helitron* by recombination or rearrangement instead of *Helitron* mediated function. We then searched the extant plant genome databases and

discovered progenitor candidate genes of *CYP72A27-Zm*. Our phylogenetic analysis of *CYP72A27-Zm* related genes from maize and grasses indicates that *CYP72A27-Zm* gene was copied from the *CYP72A26-Zm* gene by a *Helitron* around 3 mya. This *Helitron* then transposed twice or was duplicated giving rise to *CYP72A36-Zm* and *CYP72A37-Zm* around 1.5–1.6 mya. Of the four gene sequences within this *Helitron*, *CYP72A27-Zm* represents the major portion of an intact P450 gene. It contains an open reading frame for the conserved P450 motif found in P450 monooxygenase gene family. The other three genes are smaller segments of their wild type progenitors. These contain several in-frame stop codons. Reported genes captured by *Helitrons* bear varying degrees of similarity to their wild type counterparts. It is postulated that sequences that bestow advantage to the plant or advantage to transposition are preserved and thus display more similarity to their wild type progenitors. However, it is equally probable that the degree of similarity between the captured and wild type gene simply reflects the evolutionary time of capture. *CYP72A27* represents a member of a super family of P450 heme binding proteins involved in the degradation of environmental toxins and mutagens (Schuler and Werck-Reichhart 2003). Any possible transposition advantage bestowed by the *CYP72A27-Zm* gene is not apparent. Whether the presence of this gene conditions some advantage to the survival of the plant awaits investigation.

There are several mechanisms proposed for gene acquisition by *Helitrons*, however, no concrete evidence exists in support of these hypothesis due to the lack of in vivo or in vitro system to study transposition. It is postulated that the inefficient recognition of 3' palindrome sequence as a putative signal for replication termination during transposition results in the capture of flanking 3' host insertion site sequence. This hypothesis suggests that the most recent gene capture events will be represented by gene sequences located close to the 3' terminus of the element. We note that *CYP72A27-Zm*, the major intact portion of P450 gene is located upstream of three gene sequences that bear only vestige similarity to their cognate native forms. It is also proposed that replication of the *Helitron* is initiated at the conserved 3' stem-loop structure and the capture of the gene fragments occurs at the 5' end by unknown mechanisms (Poulter et al. 2003; Brunner et al. 2005).

The contiguous alignment of the immediate flanking sequence of the *Helitron* with maize ESTs bearing strong similarity to *Opie-2* retroposon indicates that the *Helitron* likely resides inside *Opie-2* retroposon. We previously reported the presence of three *Opie-2* retroposons within the boundaries of two nearly identical *Helitrons* (Gupta et al. 2005). The insertion of maize *Helitrons* into a retroelement and vice-versa have also been reported (Morgante et al. 2005). The most prolific transposable

elements, such as retroelements, whose explosive amplification has essentially doubled the size of maize genome over the past few million years, are known to target other retrotransposons for transposition and thus evade adverse effects on the host gene expression (SanMiguel et al. 1996, 1998). The data emerging from the few maize *Helitrons* characterized suggest that *Helitrons* and retroelements may target each other as landing sites in transposition. The complete sequencing of the maize genome will provide novel insight into the possibility that these elements may exploit each other for transposition.

It is postulated that movement of genes by *Helitrons* into different regions of the genome could provide functional redundancy or alternative modes of regulation to existing genes. Our EST data suggest that the *Helitron*-captured genes, *CYP72A27-Zm*, *CYP72A36-Zm* and *CYP72A37-Zm* genes are transcribed. Since *CYP72A36-Zm* and *CYP72A37-Zm* genes contain in-frame termination codons and hence do not encode wildtype proteins, one may speculate that transcripts of *CYP72A36-Zm* and *CYP72A37-Zm* interact with the wild type transcripts and regulate expression. It is also possible that truncated translation products of *CYP72A36-Zm* and *CYP72A37-Zm* may interact with the wild type protein to produce inactive hetero-multimers and hence serve as a dominant regulator of the wildtype protein.

The location of the 5' terminal ends of ESTs within the *Helitron* suggests that transcription of *CYP72A27-Zm* initiates within the element. The possibility of the *CYP72A27-Zm* transcription by a promoter located outside the element is possible albeit highly unlikely since the *Helitron* was transposed into a truncated *Opie* element oriented opposite to the *Helitron* insertion. Initiation of transcription of captured genes within *Helitrons* has been suggested previously (Brunner et al. 2005). Similarly, transcription appears to have been initiated from within the element in more than 90% of the Pack-MULEs that express captured gene fragments, (Jiang et al. 2004).

Similar to pack-MULEs, the ability of maize *Helitrons* to capture fragments of different genes and move them across the genome may have played an enormous role in exon shuffling and gene duplication. The emerging data suggest that gene movement by *Helitrons* may have substantially contributed to the massive lack of gene collinearity recently reported in modern maize inbred lines (Fu and Dooner 2002; Song and Messing 2003; Lai et al. 2005; Lal and Hannah 2005a, b). For example, Morgante et al. (2005) using oligonucleotide probes demonstrated that ~20% of the genes or gene fragments are not shared in the maize inbred lines B73 and Mo17. The majority of these changes are caused by *Helitrons* carrying fragments of cellular genes inserted in different regions of the genome. Our data provide evidence that *Helitrons* may have

played a significant role in gene amplification and evolution of the maize genome.

Acknowledgement This work immensely benefited from maize CYP72A26 and CYP72A27 genes deposited in GenBank by Dr. Mary Schuler's group, University of Illinois. We thank Dr. Schuler for her kind help and suggestions during the course of this project. The work was supported by National Science Foundation (USA) grant 0514759 to SL.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Bennetzen JM (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr Opin Genet Dev* 15:1–7
- Brunner S, Pea G, Rafalski A (2005) Origins, genetic organization and transcription of a family of non-autonomous *Helitron* elements in maize. *Plant J* 43:799–810
- Curcio MJ, Derbyshire KM (2003) The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* 4:865–877
- Dooner HK, Lal SK, Hannah LC (2006) Suggested guidelines for naming *Helitrons*. *Maize Genetics Coop Newslet* 81, www.agron.missouri.edu/mnl/81
- Elrouby N, Bureau TE (2001) A novel hybrid open reading frame formed by multiple cellular gene transductions by a plant long terminal repeat retroelement. *J Biol Chem* 276:41963–41968
- Feschotte C, Wessler SR (2001) Treasures in the attic: Rolling circle transposons discovered in eucaryotic genomes. *Proc Natl Acad Sci USA* 98:8923–8924
- Feschotte C, Jiang N, Wessler SR (2003) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3:329–341
- Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. *Proc Natl Acad Sci USA* 99:9573–9578
- Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc Natl Acad Sci USA* 93:10274–10279
- Gu X, Fu YX, Li WH (1995) Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol Biol Evol* 12:546–557
- Gupta S, Gallavotti A, Stryker GA, Schmidt RJ, Lal SK (2005) A novel class of *Helitron*-related transposable elements in maize contain portions of multiple cellular genes. *Plant Mol Biol* 57:115–127
- Jiang N, Bao ZR, Zhang XY, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431:569–573
- Jin Y-K, Bennetzen JL (1994) Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *Bs1* retroelement of maize. *Plant Cell* 6:1177–1186
- Kapitonov VV, Jurka J (2001) Rolling circle transposons in eukaryotes. *Proc Natl Acad Sci USA* 17:8714–8719
- Kapitonov VV, Jurka J (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc Natl Acad Sci USA* 11:6569–6574
- Kawasaki S, Nitasaka E (2004) Characterization of *Tpn1* family in the Japanese morning glory: *En/Spm*-related transposable elements capturing host genes. *Plant Cell Physiol* 45:933–944
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Kumar S, Tamura K, Nei N (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Lai J, Li Y, Messing J, Dooner HK (2005) Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. *Proc Natl Acad Sci USA* 102:9068–9073
- Lal SK, Hannah LC (2005a) *Helitrons* contribute to the lack of gene colinearity observed in modern maize inbreds. *Proc Natl Acad Sci USA* 102:9993–9994
- Lal SK, Hannah LC (2005b) Massive changes of the maize genome are caused by *Helitrons*. *Heredity* 95:421–422
- Lal SK, Giroux MJ, Brendel V, Vallejos E, Hannah LC (2003) The maize genome contains a *Helitron* insertion. *Plant Cell* 15:381–391
- Lal SK, Georgelis N, Hannah LC (2008) *Helitrons*: their impact on maize genome evolution and diversity. In: Hake S, Bennetzen JL (eds) *The maize handbook: domestication, genetics, and genome* (in press)
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A, 2005 Gene duplication and exon shuffling by *Helitron*-like transposons generate intraspecies diversity in maize. *Nat Genet* 37:997–1002
- Poulter RT, Goodwin TJ, Butler MI (2003) Vertebrate helitrons and other novel *Helitrons*. *Gene* 313:201–212
- Pritham EJ, Feschotte C (2007) Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis Lucifugus*. *Proc Natl Acad Sci USA* 104:1895–1900
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons in maize. *Nat Genet* 20:43–45
- SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposon in the intergenic regions of the maize genome. *Science* 274:765–768
- Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. *Ann Rev Plant Biol* 54:629–667
- Song R, Messing J (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. *Proc Natl Acad Sci USA* 100:9055–9060
- Talbert LE, Chandler VL (1988) Characterization of a highly conserved sequence related to mutator transposable elements in maize. *Mol Biol Evol* 5:519–529
- Usuka J, Brendel V (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, Zhu W, Brendel V (2000) Optimal spliced alignment of homologous cDNA to a genomic DNA template. *Bioinformatics* 16:203–211
- Vale RD, Flettereick RJ (1997) The design plan of kinesin motors. *Annu Rev Cell Dev Biol* 13:745–777
- Xu JH, Messing J (2006) Maize haplotype with a *Helitron*-amplified cytosine deaminase gene copy. *BMC Genet* 7:52
- Yang Z (1994) Estimating the pattern of nucleotide substitution. *J Mol Evol* 39:105–111
- Zhang Z, Gerstein M (2004) Large scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev* 14:328–335
- Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation, The University of Texas at Austin