

Helitrons contribute to the lack of gene colinearity observed in modern maize inbreds

Shailesh K. Lal* and L. Curtis Hannah**

*Department of Biological Sciences, Oakland University, Rochester, MI 48309-4401; and †Department of Plant Molecular and Cellular Biology, University of Florida, Gainesville, FL 32611

Until recently, it was assumed that the order of gene sequences within modern maize would be virtually invariant.

Recent discoveries have shown that gene colinearity is not always the case. Several laboratories (1–3) have found DNA regions rich in gene sequences that are present in some maize inbred lines but absent at homologous sites in other lines. This variation, termed “intraspecific violation of genetic colinearity” or “plus/minus genetic polymorphism,” was shown by Lai *et al.* (4) in a recent issue of PNAS to be caused by a newly described transposable element family termed *Helitrons*.

Lai *et al.* (4) revisited the 110-kb region of chromosome 9, which contains a number of genes, most notably the well characterized *bronze-1* locus. Earlier work by this group (1) showed that whereas the inbred McC (for McClintock) contained sequences of 10 genes in this region, only 6 of these genes were present in this region in B73, the model inbred chosen for complete DNA sequencing. In addition, hybridization experiments showed that other lines exhibited plus/minus polymorphisms for these four gene sequences that were different from McC or B73.

Lai *et al.* (4) showed that the presence of two *Helitrons* in McC and their absence in B73 totally accounts for the plus/minus variation of these four gene sequences. One *Helitron*, termed *HelA*, is 5.9 kb and contains three of the four polymorphic gene sequences described earlier (1). A second *Helitron*, *HelB*, is 2.7 kb and contains the fourth gene sequence distinguishing McC and B73. Sites of insertion of the two elements are separated by only 900 bp. A second, virtually identical copy of *HelA* was also found in the genome, albeit on a different chromosome. This second copy was also polymorphic in maize lines.

Helitrons are eukaryotic transposable elements (5–8). Because *Helitrons* share only a few invariant sequences, they remained virtually unknown until only recently. The existence of *Helitrons* was first suggested (5) from computer analysis of a class of repetitive DNA in the *Arabidopsis*, rice, and *Caenorhabditis elegans* genomes. Unlike conventional transposable elements, *Helitrons* lack

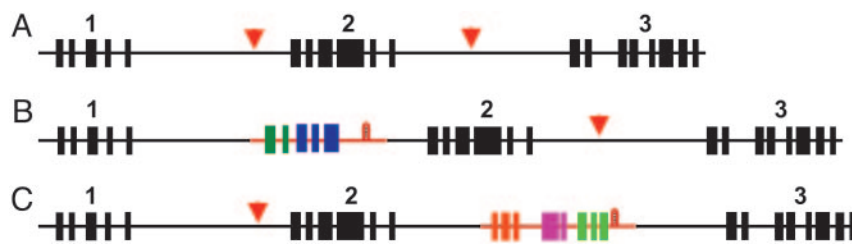


Fig. 1. Conceptual display of *Helitron*-mediated violation of gene colinearity among different maize inbred lines. Hypothetical variable haplotypes of three maize inbreds are shown. The wild-type genes are numbered, and their exons are black boxes. *Helitron* insertions are red lines, and the exons of genes captured by *Helitrons* are shown as colored boxes. The 3' palindromes of the *Helitrons* are shown as loops. The “vacant sites” in lines lacking *Helitron* insertions found in other lines are marked by red arrowheads.

terminal repeats and do not duplicate host sequences upon insertion. The only invariant sequences are a 5' TC and a 3' CTRR. All *Helitrons* so far investigated have a 10- to 16-bp palindrome near the 3' end and insert within a host dinucleotide, AT. *Helitrons* were subsequently discovered in vertebrate, white rot fungus, and mosquito genomes (6). A more concrete existence for *Helitrons* came from the finding that mutations at two maize genes, *shrunk-2* (*Sh2*) (7) and

Intraspecific violation of genetic colinearity is caused by a transposable element family termed *Helitrons*.

barren stalk (*Ba1*) (8, 9), were due to insertion of *Helitrons*. The computer-simulated autonomous *Helitron* potentially encodes helicase and nuclease/ligase. Because these proteins bear similarity to the proteins required for the rolling circle (RC) replication/transposition of some bacterial insertional sequence elements, it is widely postulated that *Helitrons* may also move by means of a similar mechanism involving replication and strand replacement (10, 11).

A most remarkable feature of *Helitrons* is their ability to capture gene sequences. These sequences contain introns, pointing to acquisition from genes

rather than mature transcripts. Although other transposable elements also incorporate gene pieces (reviewed in ref. 4), the large size of *Helitrons* allows the confiscation of segments from several genes. A 17.7-kb *Helitron*, for example, contains sequences from 12 different genes (7). Although only a few maize *Helitrons* have been sequenced, data up to now suggest that *Helitrons* are greatly enriched in gene sequences compared with the remainder of the genome. Conceptual *Helitrons* within host DNA are shown in Fig. 1.

The work of Lai *et al.* (4) provides fundamentally important insight into the mechanism underlying the infringement of gene colinearity observed in maize. However, a number of extremely important questions remain.

Can *Helitrons* capture whole genes? So far, only gene fragments have been found in *Helitrons*. Although there is no *a priori* reason to believe that whole genes cannot be contained in mobile *Helitrons*, so far, no complete genes have been found. Assuming whole gene uptake, *Helitrons* provide an intriguing mechanism for changing dosage of single genes within a single genome and creating genetic heterozygosity within completely homozygous lines. A possible role for *Helitrons* in creating the heterozygosity underlying the vigor associated with hybrids compared with inbred parents remains an intriguing possibility.

See companion article on page 9068 in issue 25 of volume 102.

**To whom correspondence should be addressed. E-mail: hannah@mail.ifas.ufl.edu.

© 2005 by The National Academy of Sciences of the USA

It is also conceivable that *Helitrons* shuffle exons and, in doing so, create novel proteins.

What are the mechanisms that underlie fragment capture and transposition of *Helitrons*? A strong case can be made for RC DNA replication because the simulated autonomous *Helitron* contains information for two enzymes (helicase and nuclease/ligase) used in this process. What is far from clear is how *Helitrons* differ so greatly in both size and sequence. Feschotte and Wessler (12) speculated that sequence divergence may arise from inefficient recognition of replication stop signals presumed to be caused by the palindrome near the 3' end. If the hairpin is not recognized, additional host sequences would be incorporated into the *Helitron*. This model predicts, then, that *Helitrons* would contain common 5' sequences but unique 3' sequences.

Although the number of *Helitrons* characterized so far is small, patterns emerging from their sequences likely bear on the origin of sequence divergence among *Helitrons*. A comparison of the *HelA* reported here with the *Helitrons* found in *Sh2* and *Ba1* shows that the elements have virtually identical terminal sequences. Sequence identity occurs in the 5' terminal 15 base pairs and the 3' terminal 30 base pairs. Captured gene pieces in the middle of the elements are totally different. A similar

pattern is observed when comparing *HelB* with a rice *Helitron*. This pattern would not be expected if sequence diversity were derived from incorporation of different host sequences in the 3' terminus of the *Helitron*.

Is there an autonomous *Helitron* element, and which enzyme functions are encoded by it?

The capture of gene sequences by site-specific recombination involving circular intermediates similar to bacterial integrons has also been proposed (9). This model accounts for common termini with unique sequences within the body of the *Helitron*. However, integrons contain an integrase gene, a gene so far not found in extant *Helitrons*.

Are *Helitrons* still mobile in modern maize? The finding of frequent plus/minus polymorphisms in *Helitron* location, the finding of virtually identical copies of *HelA* on two separate chromosomes, and the isolation of two mutants caused by *Helitrons* show that *Helitrons* have been active in maize. The precise time of movement, however, is unknown. Is there an autonomous *Helitron*

element, and, if so, which enzyme functions are encoded by it? If such an element exists, what is its distribution within maize and its wild relatives? Presently, it is impossible to determine whether the gene remnants found in a particular *Helitron* provided a function for transposition or gene capture or whether the remnant is simply a diverged fragment of a cellular gene. The precise mode of transposition and the mechanism of gene sequence acquisition by these enigmatic elements will be greatly facilitated in the future by the identification of an autonomous *Helitron* and establishment of both a genetic and a biochemical assay system with which to study their transposition. In addition, the very real possibility that sequence information for a large portion of the B73 genome will be available in the near future should aid in the identification of an autonomous *Helitron* element if that line has one.

Large-scale capture of cellular gene fragments by transposable elements is not limited to *Helitrons* or to maize. Jiang *et al.* (13) recently showed that members of the *Mutator* family of transposable elements in rice contained >1,000 gene fragments. These recent data from maize and rice add to a growing body of evidence concerning the roles that transposable elements play in duplication of genes, their movement to other sites in the genome, and possibly the structure of the genes themselves.

1. Fu, H. & Dooner, H. K. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 9573–9578.
2. Song, R. & Messing, J. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 9055–9060.
3. Brunner, S., Fengler, K., Morgante, M., Tingey, S. & Rafalski, A. (2005) *Plant Cell* **17**, 343–360.
4. Lai, J., Li, Y., Messing, J. & Dooner, H. K. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 9068–9073.
5. Kapitonov, V. V. & Jurka, J. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 8714–8719.
6. Poulter, R. T., Goodwin, T. J. & Butler, M. I. (2003) *Gene* **313**, 201–212.
7. Lal, S. K., Giroux, M. J., Brendel, V., Vallejos, C. E. & Hannah, L. C. (2003) *Plant Cell* **15**, 381–391.
8. Gallavotti, A., Zhao, Q., Kyojuka, J., Meeley, R. B., Ritter, M. K., Doebley, J. F., Pe, M. E. & Schmidt, R. J. (2004) *Nature* **432**, 630–635.
9. Gupta, S., Gallavotti, A., Stryker, G. A., Schmidt, R. J. & Lal, S. K. (2005) *Plant Mol. Biol.* **57**, 115–127.
10. Khan, S. A. (2000) *Mol. Microbiol.* **37**, 477–484.
11. Tavakoli, N., Camanducci, A., Dodd, H. M., Lett, M. C., Albiger, B. & Bennett, P. (2000) *Plasmid* **44**, 66–88.
12. Feschotte, C. & Wessler, S. R. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 8923–8924.
13. Jiang, N., Bao, Z. R., Zhang, X. Y., Eddy, S. R. & Wessler, S. R. (2004) *Nature* **431**, 569–573.