

# A cornucopia of *Helitrons* shapes the maize genome

Cédric Feschotte<sup>1</sup> and Ellen J. Pritham

Department of Biology, University of Texas, Arlington, TX 76019

**T**ransposable elements (TEs) are mobile segments of DNA first discovered in corn by Barbara McClintock in the 1940s. Her studies revealed that the movement of TEs could cause chromosomal breakage and affect phenotypes in dramatic or subtle ways. Her findings met with some disbelief because they challenged the tenet of the static genome by showing that the genome could be much more dynamic than previously appreciated. It is difficult to imagine what McClintock would have to say if she were to learn that the genome of her beloved corn is, in fact, mostly composed of TEs whose capacity to alter the genome seems limited only by the sensitivity of the techniques available to detect the alterations they engender. In this issue of PNAS (1, 2) two groups have deployed innovative computational approaches to scan the nearly complete genome sequence of maize (21) to establish a comprehensive catalog of *Helitrons*, a colorful group of TEs with a peculiar predisposition to restructure genes and genomes.

## *Helitrons* Enter the Dance

Much has been learned in the ensuing 60+ years since the discovery of the first TEs in corn. Perhaps it would come as no surprise to McClintock that many important TE discoveries were made in plants. Among recent milestones lies the discovery of an exotic group of eukaryotic TEs coined *Helitrons* (3) or rolling-circle transposons by virtue of their resemblance, both in terms of coding and structural properties, to bacterial transposons, plasmids, and viruses that replicate via a rolling-circle mechanism. Despite their abundance in some species (3, 4) and the fact that their movement has caused spontaneous mutations (5, 6), *Helitrons* were not discovered until 1999 when the first (partial) copies were computationally identified in the mustard weed *Arabidopsis thaliana* (7). It took another 2 years before the sequences of protein-encoding and presumably autonomous (i.e., self-sufficient) *Helitrons* could be unearthed from the *Arabidopsis* genome, but also from rice and nematode (3). These long (>5 kb) and potentially complete *Helitrons* all were found to encode a putative protein with rolling-circle initiator motif and PIF1-like DNA helicase domains. In addition, plant *Helitrons* were predicted to encode from one to three homologs of replication protein A

(RPA)-like ssDNA binding proteins, which are necessary for the rolling-circle replication of certain plasmids, bacteriophages, and plant geminiviruses, but are normally supplied by the host. Because of the restricted distribution of the RPA-like genes to plant *Helitrons*, it was hypothesized that they had been kidnapped from their host and retained because of their function in transposition (3). This suspicion was later reinforced by the characterization of the first maize *Helitrons*, which were found to be particularly corpulent (>10 kb) because of the capture of gene fragments from multiple chromosomal locations (5, 8–10).

## Maize *Helitrons* have transduplicated and reshuffled a staggering amount of sequences.

### Highway to *Helitrons*

*Helitrons* have been particularly recalcitrant to automated computational identification because of their lack of terminal repeats and propensity for gene capture, which creates extreme heterogeneity in size and sequence. This hurdle has now been cleared by the development of two new programs designed to recognize the peculiar terminal sequence features of *Helitrons*: HelitronFinder by the Dooner group (11) and HelSearch by Yang and Benetzen, recently benchmarked in another PNAS article (12). These programs are welcome additions to the expanding toolbox for automated annotation of TEs (13).

Using the new programs to scan the ≈2 Gb of genome sequences generated for the maize inbred line B73 (21), the two groups identified a largely overlapping set of ≈2,000 *Helitrons* predicted with high confidence (1, 2). Using this library to query the rest of the B73 sequence, many more pieces of *Helitrons* were identified (>20,000 according to ref. 2), amounting to at least 2% of the nuclear genome. The high level of fragmentation is likely an indicator of the patchy nature of the current genome assembly but also of the *Helitrons* themselves. Thus, these numbers should be

regarded as conservative and remind us that measuring the actual content of *Helitron*-derived sequences is a challenging and uncertain exercise.

The classification of *Helitrons* into traditional TE families is also problematic. Typically, members of the same TE family share high levels (>80%) of sequence identity across all or most of their length, reflecting their descent from one or a few active ancestors. However, the sequence and length heterogeneity of maize *Helitrons* coupled with our limited knowledge of the transposition mechanism made it necessary to adopt a specific set of rules for *Helitron* classification based on the most terminal regions (2, 14). Yang and Benetzen (2) considered elements as belonging to the same family when the last 30 bp are >80% identical and the same subfamily when they also share >80% sequence identity in the first 30 bp. Based on these criteria, they subdivided 1,930 intact *Helitrons* into eight families and 62 subfamilies. Du et al. (1) used a similar system of classification but with a more stringent criterion for the definition of a subfamily, thereby clustering the elements into more subfamilies (see refs. 1 and 2 for correspondence between their subfamilies). Both studies found that nearly all maize *Helitrons* (>98% in ref. 2) belong to a single (super)family called *Hip* (2) or *Hell* (1, 14) that includes several elements that have spread very efficiently, reaching several thousand copies in the B73 genome.

### Exon Shuffling Machines

The premise that *Helitrons* act as “exon shuffling machines” (15) gained support with the discovery of the first maize *Helitrons*, which were found to have transduced and often recombined exons from multiple genes leading to novel transcriptional units (5, 8–10). However, it remained unclear whether this phenomenon was a peculiarity of a subset of *Helitrons* or a mere reflection of the remarkable plasticity of the maize genome (9, 16, 17). The new studies (1, 2) provide some resolution to this question. First, they confirm that maize *Helitrons* have transduplicated and reshuffled a

Author contributions: C.F. and E.J.P. wrote the paper.

The authors declare no conflict of interest.

See companion articles on pages 19916 and 19922.

<sup>1</sup>To whom correspondence should be addressed. E-mail: cedric@uta.edu.

staggering amount of sequences, probably exceeding 20,000 gene fragments in the B73 genome (2). Of 272 *Helitrons* individually annotated by Du et al. (1), 258 (94%) carry host sequences captured from a total of 376 different genes. Although the majority of maize *Helitrons* carry one or two gene fragments, some carry exons from as many as nine different genes. Not many *Helitrons* seem to have transduced a complete gene and few have reached high copy numbers. One intriguing exception is the abundant *Hip1* (*Hell-105*) subfamily (>1,000 copies) that carries a chunk of the maize phosphatase 2C-like gene. The functional consequence of this amplification is unknown, but the captured exons display the signature of purifying selection (2).

Although most of the capture events involved the *Hip* family, the predominant group of *Helitrons* in the maize genome, many of the other families include members with transduced gene fragments. Relatives of the *Hip* family are also among the most abundant *Helitrons* in rice and sorghum. Yet, very few *Helitron*-mediated transduction events have been reported in these other grasses (12, 17). Together these observations suggest that the propensity for maize *Helitrons* to capture gene fragments has more to do with some peculiar properties of the maize genome (e.g., replication/repair fidelity) or the population genetics of the species than with the elements themselves.

### How Are Genes Captured?

In the absence of an experimental system to study the movement of *Helitrons*,

inferences of the mechanisms underlying *Helitron* gene transduction have been made by comparison with bacterial rolling-circle transposons. The latter have been shown to transduce flanking host sequences either via misrecognition of initiation sites upstream of the element's origin of replication (*ori*) or downstream through bypass of its termination (*ter*) signal (18). Although the existence and position of *ori* and *ter* remain to be determined for *Helitrons*, previous studies in maize (8), *Aspergillus nidulans* (19), and bats (4) suggested that *Helitrons* can capture flanking sequences in either direction, and that the acquisition of new terminal sequences is a potent mechanism for the evolution of new *Helitrons*. The new studies (1, 2) confirm and further these observations through several striking examples illustrating the structural plasticity of maize *Helitrons* and their ability to form new subfamilies via the accretion of new 5' and 3' termini.

A less understood aspect of the transduction mechanism is the strong bias (estimated as 14 to 1 in ref. 2) for captured gene fragments to be in the same (or "sense") orientation as the Rep/Helicase gene encoded by autonomous *Helitrons*. One explanation would be that when transcribed, exons captured in the antisense orientation are more likely to interfere with expression of the parental gene, for example, via the RNA interference pathway. Consistent with this hypothesis, Yang and Bennetzen (2) found that the few *Helitrons* containing gene fragments in antisense orientation are all of very recent origin, whereas those with fragments in sense

orientation have a broader age distribution. This pattern is consistent with natural selection acting to remove *Helitrons* carrying gene fragments that interfere with parental gene expression.

### Forecast

Although the genomes of many organisms are known to harbor substantial populations of *Helitrons* (3, 4, 12), none are known to be as recent as those of maize. Yang and Bennetzen (2) estimated that most of the elements identified in the B73 genome inserted <250,000 years ago. In addition, experiments conducted by Du et al. (1), and those published previously (5, 9, 10, 16), have shown that most *Helitron* insertions are polymorphic among maize inbred lines. These results yield two important conclusions. First, it is likely that *Helitrons* are still transposing at an appreciable frequency in some lines (see also ref. 20), making corn the species of choice to isolate and analyze the behavior of an autonomous *Helitron* in the near future. In fact, both of the new studies (1, 2) identified a few seemingly intact *Helitrons* in the B73 genome that are prime contenders as autonomous elements. Second, when one considers the sheer amount of DNA and the plethora of gene fragments carried by maize *Helitrons* it becomes evident that these transposons are responsible for much of the haplotype diversity exhibited by this model grass species (10, 16). How much of this dazzling genomic plasticity is correlated to phenotypic variation remains the burning question, and one indeed that will certainly receive further attention.

- Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic *Helitron* landscape of the maize genome. *Proc Natl Acad Sci USA* 106:19916–19921.
- Yang L, Bennetzen JL (2009) Distribution, diversity, evolution, and survival of *Helitrons* in the maize genome. *Proc Natl Acad Sci USA* 106:19922–19927.
- Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci USA* 98:8714–8719.
- 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.
- Lal SK, Giroux MJ, Brendel V, Vallejos CE, Hannah LC (2003) The maize genome contains a *Helitron* insertion. *Plant Cell* 15:381–391.
- Choi JD, Hoshino A, Park KI, Park IS, Iida S (2007) Spontaneous mutations caused by a *Helitron* transposon, *Hel-It1*, in morning glory, *Ipomoea tricolor*. *Plant J* 49:924–934.
- Surzycki SA, Belknap WR (1999) Characterization of repetitive DNA elements in *Arabidopsis*. *J Mol Evol* 48:684–691.
- Brunner S, Pea G, Rafalski A (2005) Origins, genetic organization, and transcription of a family of nonautonomous *helitron* elements in maize. *Plant J* 43:799–810.
- 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.
- Morgante M, et al. (2005) Gene duplication and exon shuffling by *helitron*-like transposons generate intraspecies diversity in maize. *Nat Genet* 37:997–1002.
- Du C, Caronna J, He L, Dooner HK (2008) Computational prediction and molecular confirmation of *Helitron* transposons in the maize genome. *BMC Genomics* 9:51.
- Yang L, Bennetzen JL (2009) Structure-based discovery and description of plant and animal *Helitrons*. *Proc Natl Acad Sci USA* 106:12832–12837.
- Feschotte C, Keswani U, Ranganathan N, Guibotsy ML, Levine D (2009) Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. *Gen Biol Evol* 1:205–220.
- Dooner HK, Hannah LC, Lal SK (2007) Suggested guidelines for naming *Helitrons*. *Maize Genetics Coop Newsletter* 81:24–25.
- Feschotte C, Wessler SR (2001) Treasures in the attic: Rolling circle transposons discovered in eukaryotic genomes. *Proc Natl Acad Sci USA* 98:8923–8924.
- Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. *Proc Natl Acad Sci USA* 103:17644–17649.
- Sweredoski M, DeRose-Wilson L, Gaut BS (2008) A comparative computational analysis of nonautonomous *helitron* elements between maize and rice. *BMC Genomics* 9:467.
- Toleman MA, Bennett PM, Walsh TR (2006) ISCR elements: Novel gene-capturing systems of the 21st century? *Microbiol Mol Biol Rev* 70:296–316.
- Cultrone A, Dominguez YR, Drevet C, Scazzocchio C, Fernandez-Martin R (2007) The tightly regulated promoter of the *xanA* gene of *Aspergillus nidulans* is included in a *helitron*. *Mol Microbiol* 63:1577–1587.
- Li Y, Dooner HK (2009) Excision of *Helitron* transposons in maize. *Genetics* 182:399–402.
- Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science*, 10.1126/science.1178534.