Chapter 11
Helitron Proliferation and Gene-Fragment Capture

Yubin Li and Hugo K. Dooner

Abstract *Helitrons* stand out as rare transposons discovered by bioinformatic, rather than genetic, studies. Although they comprise an ancient superfamily of transposons found in plants, animals, and fungi, it is in plants where they have been studied most extensively. Well-annotated plant genomes contain increasingly higher numbers of identified *Helitrons*, including putative autonomous elements and nonautonomous elements with and without gene fragments. The molecular structure of the autonomous *Helitron* and the postulated rolling circle mode of transposition remain hypothetical, and recent evidence suggests that *Helitrons* may transpose by both copy-and-paste and cut-and-paste mechanisms. Two *Helitron* properties, in particular, have caught the imagination of biologists: their ability to undergo sudden bursts of transposition and their ability to capture fragments from different genes to make chimeric transcripts. In this chapter, we provide an overview of what we have learned in the past decade about the biology of these intriguing, newly discovered plant genome residents.

Keywords *Helitrons* • Transposons • Plants

11.1 Introduction

Transposable elements (TEs) are DNA fragments that can move from one site of the genome to another. Though ubiquitous in nature, they were first discovered in maize more than 60 years ago (McClintock 1947). This eventual Nobel-Prize-winning

Y. Li
Waksman Institute, Rutgers University, Piscataway NJ, 08854, USA
e-mail: yubin@waksman.rutgers.edu

H.K. Dooner (✉)
Waksman Institute, Rutgers University, Piscataway NJ, 08854, USA

Department of Plant Biology, Rutgers University, New Brunswick, NJ 08901, USA
e-mail: dooner@waksman.rutgers.edu

M.-A. Grandbastien and J.M. Casacuberta (eds.), *Plant Transposable Elements*, Topics in Current Genetics 24, DOI 10.1007/978-3-642-31842-9_11,

discovery began to be acknowledged broadly only three decades later and gained increasingly wider appreciation in the "omics" era (Craig et al. 2002). Today, TEs are considered to have played an intrinsic role in genome structure evolution through the multiple chromosome rearrangements that are brought about by the chromosome cutting properties noted by McClintock (1952). TEs have been proposed as a major driving force in the process of gene creation by providing the raw material needed for the evolution of new gene functions (Dooner and Weil 2012; Feschotte and Pritham 2007) and have turned out to be the major component of most sequenced eukaryotic genomes (Craig et al. 2002).

At the turn of the twenty-first century, the known classes of TEs (Feschotte et al. 2002) were expanded to include the newly hypothesized *Helitron* transposable elements. Unlike Class I elements (retrotransposons) that transpose through RNA, Class II elements (DNA transposons) transpose through DNA. *Helitrons* were postulated to transpose via a hypothetical rolling circle (RC) replication mechanism (Kapitonov and Jurka 2001) and, therefore, fall into the latter class. A more recent classification of eukaryotic transposons places them under a special Subclass 2 among DNA transposons (Wicker et al. 2007). In the past decade, a considerable effort has been made to better understand these elusive TEs from all different angles. Our goal in this chapter is to summarize our current knowledge about these DNA transposons in the plant kingdom and to provide a personal view of further explorations in this emerging field.

11.2 Discovery of *Helitrons*

Shortly before their discovery as unique eukaryotic transposons, *Helitrons* had been described as repetitive sequences in *Arabidopsis thaliana*, one of the three genomes analyzed by Kapitonov and Jurka (2001) in their seminal paper. The first such repeat detected was *Aie* (Arabidopsis insertion element), a 527-bp element insertion present downstream of the polyadenylation site of *AtRAD51* in the Columbia ecotype but absent in its Landsberg *erecta* counterpart (Doutriaux et al. 1998). *Aie* is AT-rich, contains no ORFs, has a stem-and-loop sequence on the 3′ side (5 unpaired bases in a 21-bp stem, with a 4-bp loop), and shows some short duplications around the insertion site. Because it lacked terminal inverted repeats (TIRs), *Aie* was taken to be a remnant of an imperfect transposition event, an interpretation supported by its multicopy presence in the two ecotypes.

Due to their abundance in the genome, elements closely related to *Aie* were readily uncovered in subsequent computational analyses of *Arabidopsis* repetitive sequences. *AthE1* was the most abundant class of repetitive elements in the *A. thaliana* 1998 sequence database (Surzycki and Belknap 1999). Although they could be as long as 2 kb, these elements lacked any detectable coding capacity for known transposases. While the 5′ and 3′ ends of *AthE1* family members were highly conserved, they did not represent either inverted or direct repeats. Direct repeats flanking transposons, also known as target site duplications (TSD), are a common feature of

retrotransposons and DNA transposons. Their absence in *AthE1* elements suggested that these elements differed from most other known transposons in being unable to recombine into the genome by introducing staggered cuts in the target DNA.

In a comprehensive analysis of potential transposon sequences in chromosome 2 of *Arabidopsis*, sequences resembling *AthE* were found to make up 1.1 % of the chromosome. No detectable TSDs or TIRs flanked these unusual repeats, which were named ATREP1-10 and classified as ten families of nonautonomous DNA transposons (Kapitonov and Jurka 1999). Another analysis of transposon diversity in a much larger *Arabidopsis* dataset (≈17.2 Mb) grouped 179 *AthE*-like or *ATREP*-like elements into seven families based on common structural features and identified them as members of a novel superfamily of transposons, named *Basho*, that moved by an unknown transposition mechanism (Le et al. 2000). A *Basho*-like group was also identified in maize, supporting the concept of a new plant transposon superfamily. Completion of the whole genome sequence of *Arabidopsis* (Arabidopsis Genome Initiative 2000) revealed the existence of 1,265 *Basho* elements. In contrast with the class I elements that primarily occupy the centromere, but consistent with other class II transposons, *Basho* elements predominate on the periphery of pericentromeric domains. Novel elements resembling the structurally unusual *Basho* elements were also found in rice, suggesting a wide distribution of these elements in plants (Turcotte et al. 2001). Similar to *Basho* elements in *Arabidopsis*, the rice elements are small (<2 kb), lack coding capacities, TSDs or TIR, and are highly conserved at both termini. The big outstanding question after these studies was: by what mechanism does this new superfamily of transposons multiply and transpose in the host genome?

In 2001, this question was answered hypothetically when Kapitonov and Jurka (2001) carried out an in silico reconstruction of putative autonomous transposons from inactive copies accumulated in the three genomes analyzed, *Arabidopsis thaliana*, *Caenorhabditis elegans*, and *Oryza sativa*. Deletions, insertions, and premature stop codons were removed from the consensus sequences of the transposons by computational approaches, in a reconstruction process reminiscent of that of *Sleeping Beauty* (Ivics et al. 1997). Finally, rolling circle (RC) replication, a transposition mechanism until then restricted to prokaryotes, was proposed to explain movement of this previously unknown category of eukaryotic DNA transposons. The new elements were designated *Helitrons* because the protein encoded by the putative autonomous elements had a conserved DNA helicase domain.

11.3 Genomics of *Helitrons*

11.3.1 Molecular Structure of Putatively Autonomous and Nonautonomous Helitrons

Helitrons have been found in every plant genome where they have been carefully looked for (Table 11.1). As a consequence of their in silico detection, the majority of *Helitrons* identified in a given species share distinct structural features with other

Table 11.1 Dynamic distribution of *Helitron* transposons in sequenced plant genomes

Organism	Genome size (Mb)	Family no.	Putative autonomous	Occurrence no.	Genome fraction (%)	References
Arabidopsis thaliana	115	NA		1,265	NA	Arabidopsis Genome Initiative (2000)
	115	4	+	910	~2.0	Kapitonov and Jurka (2001)
	115	10		1,242	1.30	Yang and Bennetzen (2009b)
	115	NA		3,437	1.85	Hollister et al. (2011)
	119	34		12,947	6.72	Ahmed et al. (2011)
Arabidopsis lyrata	206.7	NA		10,452	2.64	Hollister et al. (2011)
Brachypodium distachyon	272	48		120	0.18	International Brachypodium Initiative (2010)
Brassica rapa	284	NA		6,214	0.60	Wang et al. (2011)
Glycine max	975	NA		7,128	0.53	Schmutz et al. (2010)
				82	NA	Du et al. (2010)
Medicago truncatula	243	10	+	1,386	1.29	Yang and Bennetzen (2009b)
Oryza sativa var. japonica	389	NA		552	NA	Sweredoski et al. (2008)
		NA		3,037	0.33	Paterson et al. (2009)
		23	+	6,947	2.09	Yang and Bennetzen (2009b)
Oryza sativa var. indica		NA		604	NA	Sweredoski et al. (2008)
Physcomitrella patens	480	1	+	19	0.12	Rensing et al. (2008)
Populus trichocarpa	485	2		NA	0.06	Tuskan et al. (2006)
Selaginella moellendorffii	213	4	+	5,394	1.57	Banks et al. (2011)
Sorghum bicolor	748	1		1,017	0.81	Paterson et al. (2009)
		11	+	4875	3.00	Yang and Bennetzen (2009b)
Vitis vinifera	487	NA		109	0.01	Jaillon et al. (2007)
Zea mays	2,400	29[a]		2,791	2.00	Du et al. (2009)
	2,050	8		1,930	2.20	Yang and Bennetzen (2009a)

NA not available

[a]Family number of genic *Helitrons*

elements in the same species and in closely related species. The putative autonomous *Helitrons* reconstructed from nonautonomous ones in *Arabidopsis thaliana* (*Helitron1* and *Helitron2*) and *Caenorhabditis elegans* (*Helitron1_CE*) encode a large protein denominated RepHel that contains a Rep domain homologous to RC replication initiators and a Hel domain homologous to DNA helicases (Kapitonov and Jurka 2001). Because the predicted RepHel proteins share motifs with the transposases of bacterial RC transposons, *Helitrons* were postulated to transpose by RC replication. The enzymatic core of the ~100-aa Rep domain contains three motifs that are conserved in a wide diversity of eukaryotes (Feschotte and Pritham 2007; Kapitonov and Jurka 2007). The larger, ~400-aa Hel domain contains eight universally conserved motifs in all putative autonomous *Helitrons* (Fig. 11.1a). Examples of these conserved motifs are shown in Fig. 11.1d. Conservation of the RepHel protein has been used as the criterion to identify hypothetical autonomous *Helitrons* in all plant host genomes (Table 11.1).

Shorter nonautonomous *Helitrons* are far more abundant and correspond to the non-TIR-, non-TSD-containing highly repetitive sequences that were noted earlier in *Arabidopsis* and rice. They have been grouped into multiple families based on the degree of sequence conservation at both 5′ and 3′ termini (Fig. 11.1b). Most of these elements are smaller than 2 kb and encode no detectable proteins. Longer elements with extra protein-coding capacity (Fig. 11.1c) occur in some species. For example, in *Arabidopsis* and rice, the putative autonomous *Helitrons* also encode subunits of RPA70, a single-stranded-DNA-binding protein. These are absent in *C. elegans*, making it unlikely that they are part of the transposition machinery (Kapitonov and Jurka 2001). Though RPA-like proteins have also been identified in some animal *Helitrons* (Feschotte and Pritham 2007; Kapitonov and Jurka 2007), their exact function remains unknown.

11.3.2 Biological and Computational Identification of Helitrons

Among the dozens of known eukaryotic DNA transposons (Feschotte and Pritham 2007; Kapitonov and Jurka 2008; Wicker et al. 2007), *Helitrons* stand out as a rare example of TEs discovered purely by computational, rather than genetic, studies. Though only recently identified, *Helitrons* are an ancient superfamily of eukaryotic DNA transposons, as evidenced by their cross-kingdom presence in plants (Table 11.1), fungi (Galagan et al. 2005), and animals (Cocca et al. 2011; Kapitonov and Jurka 2001; Pritham and Feschotte 2007). *Helitrons* are the only eukaryotic transposons that lack TIRs, do not generate TSDs upon integration in the host genome, and do not encode any known transposases. Furthermore, until their computational discovery, none had been found to be the causative agent of a mutation. These unusual features delayed their discovery, although *Helitrons* resemble other eukaryotic DNA transposons in terms of their impact on the host genome. Following their discovery, *Helitrons* have been identified by both biological and computational approaches.

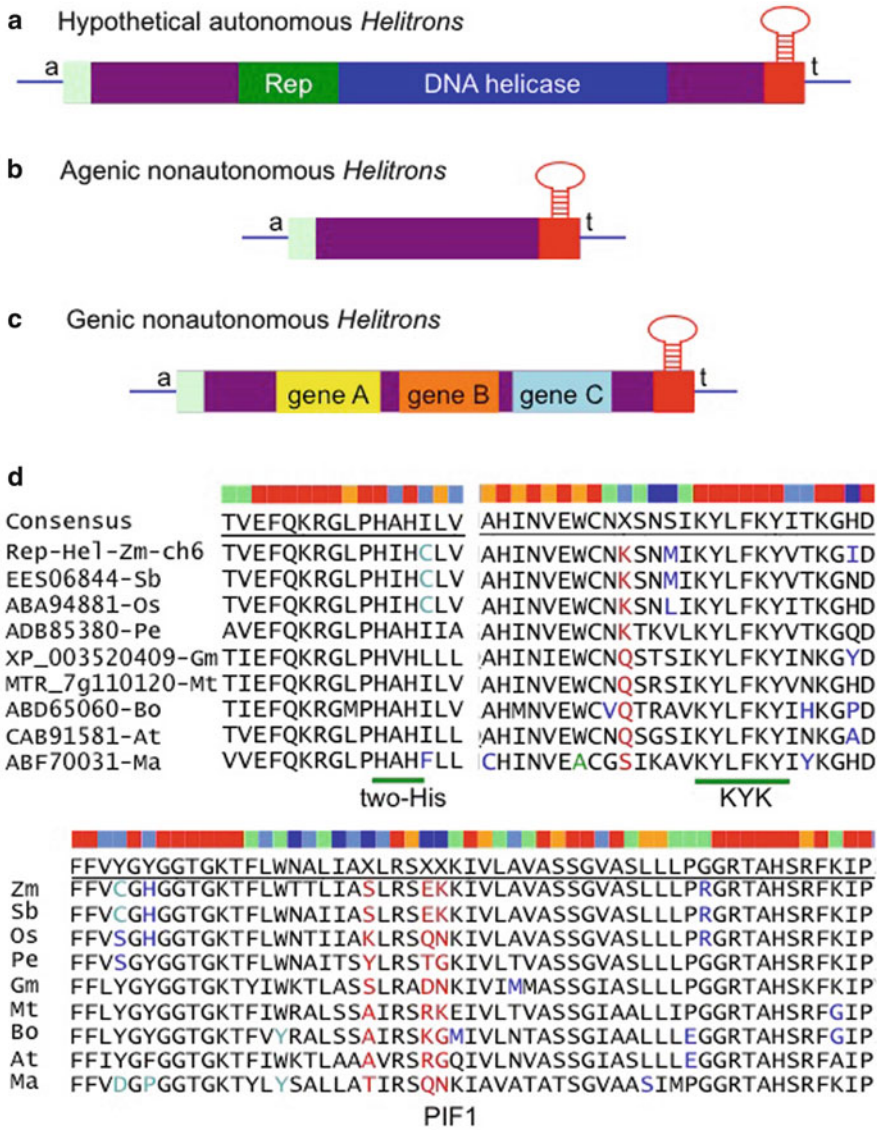

Fig. 11.1 Generic structure of identified *Helitrons* in different eukaryotes. (**a**) Hypothetical autonomous *Helitron* with coding capacity for a RepHel protein. Rep (Replication motifs are in *green*, and Hel (Helicase motifs are in *blue*. The conserved 5′ TC terminus is shown in *light green*. The conserved 3′ CTRR terminus is shown in *red*, with a stem-loop structure formed from a palindromic sequence in the 3′ subterminal region. The insertion is targeted to an AT dinucleotide shown in lowercase above a *blue line* representing the flanking sequence. The vast majority of *Helitrons* are nonautonomous elements with similar terminal structures as the autonomous copies. (**b**) Agenic nonautonomous *Helitrons* lack any known coding capacity. (**c**) Genic nonautonomous *Helitrons* carry fragments from a variable number of genes in the host genome (*yellow*, *orange*, and *light blue boxes*). (**d**) Multiple alignments of the conserved motifs of Rep domain (two-His and KYK and PIF1 helicase domain) in plant *Helitrons*. At, *Arabidopsis thaliana*; Bo, *Brassica oleracea*; Gm, *Glycine max*; Ma, *Musa acuminata*; Mt, *Medicago truncatula*; Os, *Oryza sativa*; Pe, *Phyllostachys edulis*; Sb, *Sorghum bicolor*; Zm, *Zea mays*

11.3.2.1 Biological Identification of *Helitrons*

Helitrons have been detected biologically in only a handful of cases, either as insertional mutagens causing spontaneous mutations (Table 11.2) or as colinearity disruptors contributing to haplotypic diversity within a species.

Molecular characterization of the spontaneous *sh2-7057* mutant allele in maize (Lal et al. 2003) revealed that the mutation carried a large *Helitron* insertion in the 11th intron of the *sh2* gene. This was the first case to demonstrate the mutagenicity of *Helitron* transposons. Though the insertion in this mutant was larger than 12 kb, it lacked coding capacity for known transposases and, instead, carried several gene fragments, including four exons with similarity to a plant DEAD box RNA helicase.

The strong terminal sequence similarity of the insertion in the spontaneous mutation *ba1-ref* (*barren stalk-1*) with the *Helitron* transposon in *sh2-7057* led to the realization that this classical mutation, identified more than three quarters of a century ago, had been caused by a *Helitron* insertion. In contrast to the insertion in *sh2-7057*, the 6.5 kb *Helitron* element in *ba1-ref* inserted in the proximal promoter region of the *ba1* gene (Gupta et al. 2005). Though the 6.5-kb insertion also carried multiple pseudogene fragments, these differed from those in the *Helitron* transposon of *sh2-7057*. The conserved 5′ and 3′ termini of these *Helitrons* were found to be repetitive in the maize genome, suggesting that they play an important role in *Helitron* amplification.

More strikingly, three independent *ts4* mutations, which develop carpels in the florets of the tassel, were found to carry *Helitron* insertions in the promoter of the *zma-MIR172e* gene (Chuck et al. 2007). These mutations arose at different times in different genetic backgrounds. Since only the ends of the insertions were sequenced, it is not possible to speculate on the relationships among these elements. However, the similarity in size between the insertions in *ts4-TP* and *ba1-ref* (~6 kb) suggests that the former may also carry gene fragments.

Mutations caused by *Helitron* insertions have been identified in other plant genomes, as well (Table 11.2). *Hel-It1*, the first mutagenic *Helitron* described in dicots, interrupts the anthocyanin pigmentation gene *DFR-B* in the *pearly-s* mutant of *Ipomoea tricolor* (Choi et al. 2007). This 11.5-kb *Helitron* shows the structure predicted for a plant autonomous element, with conserved 5′ and 3′ termini and genes for Rep/Hel and RPA proteins. A frameshift mutation in the former and a nonsense mutation in the latter would render this element nonautonomous, but several related elements are found in the *Ipomoea* genome. In fact, *RPA* transcripts not containing the nonsense mutation of *Hel-It1* were detected in the *pearly-s* mutant and were proposed to originate from a hypothetical autonomous element present in that line.

The 3′-UTR of genes appeared to be an underrepresented target for *Helitron* insertion until a recent study on the S-RNase-based gametophytic self-incompatibility system in the tetraploid sour cherry (*Prunus cerasus*). A 306-bp nonautonomous *Helitron* element was identified 38 bp downstream of the stop codon of the *SFB* gene in four nonfunctional (self-compatible) S_{36} variants (Tsukamoto et al. 2010). The vast majority of *SFB* transcripts in S_{36} do not have

Table 11.2 Characterized variants resulting from *Helitron* insertions

Variant name	Targeted gene	Species	Gene function	Insertion site	References
sh2-7527	*shrunken-2*	*Zea mays*	Large subunit of the tetrameric maize endersperm ADP-glucose pyrophosphorylase	Eleventh intron	Lal et al. (2003)
ba1-ref	*barren stalk1*	*Zea mays*	An atypical bHLH transcription factor that affects every axillary meristem	Proximal promoter	Gupta et al. (2005)
Hel-ts4-TP	*tassel seed4*	*Zea mays*	microRNA 172e involved in sex determination and meristem cell fate	TATA box	Chuck et al. (2007)
Hel-ts4-A	*tassel seed4*	*Zea mays*	microRNA 172e involved in sex determination and meristem cell fate	TATA box	Chuck et al. (2007)
Hel-ts4-ref	*tassel seed4*	*Zea mays*	microRNA 172e involved in sex determination and meristem cell fate	883 bp upstream	Chuck et al. (2007)
AtREP2	MEDEA	*Arabidopsis thaliana*	SET domain protein of polycomb group	3,809 bp upstream	Spillane et al. (2004)
Hel-It1	DFR-B	*Ipomoea tricolor*	Dihydroflavonol 4-reductase for anthocyanin biosynthesis	Fifth intron	Choi et al. (2007)
Helitron-Os	TnpA	*Oryza sativa cv. japonica*	CACTA element transposase	Seventh intron	Greco et al. (2005)
Helitron-Pc	SFB	*Prunus cerasus*	Pollen self-incompatibility locus	3-′UTR	Tsukamoto et al. (2010)

a poly (A) tail, suggesting that the presence of the *Helitron* element interferes with the polyadenylation process. *Helitron* elements have also been found associated with certain *S* haplotypes in the self-compatible species *Arabidopsis thaliana* (Liu et al. 2007; Sherman-Broyles et al. 2007), raising the intriguing prospect that they may have played a widespread role in the evolution of self-compatibility. However, further studies are needed to establish conclusively that the *Helitron* insertion was the real cause of the loss of function of the S_{36} variants in sour cherry.

Genome components other than genes, such as DNA transposons, can also be targeted by *Helitrons*. In *OsES1*, a rice homolog of the maize *En*/*Spm* transposon, a 1,280-bp nonautonomous *Helitron* transposon, is located in the seventh intron of the gene encoding the TnpA transposase (Greco et al. 2005). The *Helitron* insertion seems to induce alternative splicing, as do many other transposon insertions in transcribed regions (Dooner and Weil 2012). Thus, *Helitrons* may play a role in the regulation of the transpositional activity of *CACTA* elements, the most abundant superfamily of DNA transposons in rice (Paterson et al. 2009).

Because many maize *Helitrons* carry segments of multiple genes, they have been identified much more frequently as disruptors of genetic colinearity among different maize inbred lines (Brunner et al. 2005a, b; Fu and Dooner 2002; Lai et al. 2005; Morgante et al. 2005; Song and Messing 2003; Wang and Dooner 2006). The so-called "intraspecific violation of genetic colinearity" (Fu and Dooner 2002) or "plus–minus variation" (Lai et al. 2005) resulting from *Helitron* insertions in maize led to community efforts to achieve a more detailed and precise identification and annotation of *Helitrons* (Du et al. 2008, 2009; Yang and Bennetzen 2009a). This effort was essential to a proper annotation of the actual gene content in the maize genome (Schnable et al. 2009) because of the gene-fragment-rich property of the widely prevalent nonautonomous elements (Lal et al. 2009a).

Recently, a maize-type of *Helitron* transposon was discovered in the Pooideae grass *Lolium perenne* (perennial ryegrass). Large (~7.5 kb) *Helitron* elements were identified that had trapped fragments, including exons and introns, from three genes: *GIGANTEA* (*GI*), succinate dehydrogenase, and ribosomal protein S7 (Langdon et al. 2009). All three fragmented genes shared the same transcription orientation as the *Helitron* elements. Highly similar *Helitrons* were detected in the closely related grass species *Festuca pratensis* (meadow fescue), indicating a likely common ancestral origin of these elements.

11.3.2.2 Computational Identification of *Helitrons* in Sequenced Organisms

The vast majority of *Helitrons* were identified from in silico studies of sequenced genomes either manually or via investigator-designed ad hoc mining programs, such as *DomainOrganizer* (Tempel et al. 2006), *HelitronFinder* (Du et al. 2008, 2009), *HelSearch* (Yang and Bennetzen 2009b), and *Helitron_scan* (Feschotte et al. 2009). The contribution of *Helitrons* to plant genomes varies widely, from none to as high as ~7 %. However, determining an exact figure for the *Helitron* content of any given host genome is chancy. Due to the extremely limited sequence

conservation among *Helitrons*, it is not surprising to find quite different figures in updated versions of the same genome sequence (e.g., Du et al. 2010; Schmutz et al. 2010).

The published programs for automated computational identification and classification of *Helitrons* utilize either a homology-based or a structure-based approach. The latter approach (Du et al. 2008; Yang and Bennetzen 2009b) has been applied only recently in the analysis of whole genomes (Du et al. 2009, 2010; Yang and Bennetzen 2009a).

Initially, the homology-based approach was used to compare sequences at both the nucleotide and amino acid levels, as demonstrated by Kapitonov and Jurka (2001) in their original paper. *Helitron*-like transposons in rice were classified as *Helitrons* based on their capacity to code for proteins homologous to Rep/helicase and RPA (Kapitonov and Jurka 2001) and their shared structure hallmarks with *Arabidopsis Helitrons* (AT insertion site, 5′-TC, and 3′-CTRR and the 15- to 20-nucleotide palindrome close to the 3′-end). In an analogous approach, 21 *Helitron* elements were identified in the model legume *Lotus japonicus* by using as queries the RC motif and domain-5 of the RepHelicase from Arabidopsis *Helitrons*. Altogether, *Helitron* elements made up 0.4 % of the 32.4 Mb examined sequences (Holligan et al. 2006).

Novel *Helitrons* were also identified by nucleotide similarity to whole *Helitron* elements or to just the termini (Du et al. 2008, 2009; Kapitonov and Jurka 2001; Sweredoski et al. 2008; Tempel et al. 2007; Yang and Bennetzen 2009a, b). Other prevalent criteria implemented in genome-wide annotations of *Helitron* transposons include nonallelic locations in a given host genome and presence/absence of polymorphisms revealed from vertical comparison of colinear regions in closely related genomes (Wicker et al. 2010).

In addition to the two model plant genomes where *Helitrons* were originally identified, *Helitrons* have been detected in many other flowering and nonflowering plants. Paralleling the 20-fold variation in genome size, *Helitron* content varies from 0.01 % in grape to 6.72 % in the latest annotation of the *Arabidopsis thaliana* genome (Table 11.1). The estimated contribution of *Helitron* elements to a particular host genome also varies in different databases analyzed by different researchers, as seen *Arabidopsis thaliana*, rice, sorghum, and soybean.

Helitrons are poorly conserved among species, even of the same genus; this has made it hard to determine their presence systematically. Nevertheless, comparisons of the *Helitron* content of closely related species have been carried out in *Arabidopsis* and rice. The former involved the whole genomes of *A. thaliana* and *A. lyrata* (Hollister et al. 2011) and the latter, the partial genomes of 13 *Oryza* species (Gill et al. 2010).

As shown in a recent study on TE evolutionary dynamics in *Arabidopsis* employing the powerful transposon display method, *Basho Helitrons* were amplifiable in *A. thaliana* but were apparently absent from *A. lyrata*. This led to the suggestion of a recent burst of *Basho* insertions specifically within *A. thaliana* (Lockton and Gaut 2010). However, a subsequent sequence annotation effort revealed that *Helitrons* are actually the most abundant TEs in the fully sequenced *A. lyrata* genome (Hollister et al. 2011).

In an attempt to examine the relative abundance and distribution of TE classes across the genus *Oryza*, DNA transposons were identified by homology-based searches of BAC-end sequences from 13 species representing 8–17 % of each of the ten *Oryza* genome types. The *Helitron* content in the genus was found to vary greatly, from 0.29 % in *O. australiensis* to 3.15 % in *O. glaberrima* (Gill et al. 2010).

The identification of *Helitrons* from newly sequenced genomes remains a challenging endeavor despite the availability of several refined programs for detecting them. As shown in Table 11.1, *Helitron*-related sequences make up as much as 1.6 % of the *Selaginella* genome (Banks et al. 2011), but less than 0.2 % of the *Brachypodium* (International Brachypodium Initiative 2010) and *Physcomitrella* (Rensing et al. 2008) genomes. The lesson learned from other genomes, such as sorghum, suggests that the *Helitron* content of the latter two genomes will increase upon future careful annotation.

Glimpses of ongoing sequencing projects reveal that *Helitrons* are major components of some other plant genomes, as they are in sequenced model genomes. For example, *Helitron* transposons constitute ~1 % of 1.2 Mb of sequences from the tetraploid moso bamboo (*Phyllostachys pubescens* E. Mazel ex H. de Leh.) (Gui et al. 2010). In wheat (*Triticum aestivum*), 3,222 TEs have been annotated in 18.2 Mb of sequence from chromosome 3B. Only five families of agenic nonautonomous *Helitrons* were identified, representing just 0.07 % of the genomic sample sequences, in contrast to the 81.4 % contribution from all other TEs (Choulet et al. 2010). The only *Helitron* found so far in barley (Scherrer et al. 2005) is present in about 20–30 copies in the genome, based on 574 Mb of high-throughput sequences representing about 10 % of a genome equivalent (Wicker et al. 2008). Very recently, a putative *Helitron* sequence was first reported in sunflower and its insertion was dated to 1.14 million years ago (Buti et al. 2011).

In spite of the ever-growing numbers of identified *Helitrons* in newly sequenced genomes, a much more careful characterization of *Helitron* composition is necessary for sequenced plant genomes where *Helitrons* have not been yet identified, such as *Carica papaya* (Ming et al. 2008), *Cucumis sativus* (Huang et al. 2009), and *Solanum tuberosum* (The Potato Genome Sequencing Consortium 2011). Given the ubiquitous presence of these elements in all carefully annotated plant genomes, *Helitron*-free plant genomes are unlikely to exist.

11.3.3 Coding Capacity

The structure of the hypothetical autonomous *Helitron* proposed by Kapitonov and Jurka (2001) is fairly sound since elements with a similar structure continue to be found in an increasing number of genomes (Choi et al. 2007; Morgante et al. 2005). However, all of the *Helitrons* identified so far are nonautonomous and, oftentimes, bear gene fragments coding for proteins other than the REP-HEL transposase proposed for the RC transposition of *Helitrons* (Brunner et al. 2005a, b; Gupta et al. 2005; Lai et al. 2005; Lal et al. 2003; Morgante et al. 2005; Wang and Dooner 2006; Xu and Messing 2006).

In maize, two research groups have scanned the nearly complete genome sequence using similar computational approaches (Du et al. 2009; Yang and Bennetzen 2009a) and concluded that the majority of the ~2,000 genic *Helitrons* identified carried fragments from genes located in different chromosomes, with a few exceptions coming from neighboring genes. The tendency of *Helitrons* to gene-fragment capture seen in maize may be not a general property of plant *Helitrons*. For instance, in *A. thaliana*, very few *Helitron* families were found to have acquired gene fragments (Hollister and Gaut 2007; Yang and Bennetzen 2009b). A similar low propensity to capture genes was found among *Helitrons* from rice, sorghum, and *Medicago* (Yang and Bennetzen 2009b).

As is the case with most other transposon superfamilies (Levin and Moran 2011), small RNAs generated from endogenous *Helitron* sequences have the potential to inhibit TE mobility through the posttranscriptional degradation of transposon mRNA. As recently reported in *Physcomitrella patens*, 6 % of the nucleotides within 48 23-nucleotide RNA loci overlapped with regions similar to *Helitron* elements, which make up just 0.12 % of the genome (Cho et al. 2008).

11.3.4 Target Preference

The insertion site preference of *Helitron* transposons has been analyzed at the nucleotide level (target site sequence specificity), gene level (coding capacity of target sequence), and genome level (chromosomal distribution).

Plant *Helitrons* insert almost invariably in a 5′-AT-3′ dinucleotide (Brunner et al. 2005a, b; Choi et al. 2007; Gupta et al. 2005; Kapitonov and Jurka 2001; Lai et al. 2005; Lal et al. 2003; Morgante et al. 2005; Wang and Dooner 2006; Xu and Messing 2006) and, exceptionally, in a 5′-NT-3′ dinucleotide (Du et al. 2008, 2009; Morgante et al. 2005; Yang and Bennetzen 2009a). In addition, plant *Helitron* insertion sites are notably AT-enriched on either side of the insertion (Du et al. 2009; Yang and Bennetzen 2009a).

The discovery over the last decade that *Helitron* insertions have been the cause of spontaneous mutations in several plant species would suggest that *Helitrons* target genic regions (see Table 11.2), at least in these host genomes. Supporting this inference, maize *Helitrons* were found to be most abundant in gene-rich regions across the genome (Du et al. 2009; Yang and Bennetzen 2009a). However, this may not be a general pattern in plants.

In *Arabidopsis*, for example, *Helitrons* are enriched in gene-poor pericentromeric regions (Yang and Bennetzen 2009b), thus showing a pattern opposite to that of other DNA transposons, which are frequently associated with gene-rich regions. However, in a different study that compared the proximity of transposons of different ages to genes in *A. thaliana*, *Helitrons,* and other recently active TE families, such as MITEs, tended to be closer to genes than ancient families, such as CACTA-like elements (Hollister and Gaut 2009). Moreover, nonautonomous *Helitrons*, many as small as MITEs, were unmethylated in higher proportions than most other TE families.

These observations were explained by a model in which host silencing of TEs near genes has deleterious effects on neighboring gene expression, resulting in the preferential loss of methylated TEs from gene-rich chromosomal regions.

In rice, *Helitron* elements are more scattered along the chromosomes and not enriched in all pericentromeric regions (Yang and Bennetzen 2009b). As with other TEs, the distribution of *Helitrons* in present-day genomes probably reflects a combination of factors, such as continued mobility, insertion specificity, purifying selection against insertion in genes, and rates of DNA removal in gene-poor heterochromatic regions.

11.3.5 Differential Amplification and Contribution to Host Genome

The variable patterns of *Helitron* accumulation in sequenced plant genomes suggest different dynamics of *Helitron* proliferation across species and differential contributions to the present structure of their host genomes.

Helitrons make up a wide fraction of the plant genomes sequenced so far, from barely detectable to as much as 1/16 (Table 11.1). As has been well documented, TE proliferation and polyploidization are the two major processes that increase plant genome size (Bennetzen 2005). *Cornucopious*, the most abundant *Helitron* transposon subfamily in maize, consists of thousands of copies of ~1-kb agenic elements with variable sequence identity to the consensus (Du et al. 2009). These relatively small maize *Helitrons* may be actively transposing after a recent escape from transposition suppression, like the *mPing* MITEs suddenly amplified during rice domestication (Naito et al. 2006), whereas the amplification of the vast majority of *Helitron* families in maize, rice, and *Sorghum* peaked about 0.25 million years ago (Yang and Bennetzen 2009a).

In the recent annotation of the *A. thaliana* genome (Ahmed et al. 2011), *Helitron*-related sequences made up 6.7 % of the genome, more than the sum of all other DNA transposons (Table 11.1). In agreement with earlier results (Hollister and Gaut 2009), elements from the *Helitron* and *Tc1*/mariner superfamilies had the highest proportion of unmethylated sequences, whereas those from the *Gypsy* and *CACTA* superfamilies had the lowest.

As with *Helitron* content, different numbers of *Helitron* families have been identified the same organism (Table 11.1). In general, *Helitrons* with a smaller size tend to be amplified to a high degree (Ahmed et al. 2011; Du et al. 2009; Hollister and Gaut 2007). And, as noted in *Arabidopsis* and maize, longer *Helitrons* are less likely to persist in the genome (Hollister and Gaut 2007; Yang and Bennetzen 2009a), presumably because they are selected against in order to avoid the deleterious effects of inter *Helitron* ectopic recombination. However, other explanations may be possible because no recombination was detected within the heavily methylated gene fragments borne on maize *Helitrons* in a large-scale experiment specifically designed for that purpose (He and Dooner 2009).

In addition to their effect on genome size through massive amplification of agenic families, *Helitrons* contribute to haplotype variability through transposition and chromosome rearrangements (Ahmed et al. 2011; Brunner et al. 2005a; Lai et al. 2005; Morgante et al. 2005; Wang and Dooner 2006). The mechanism of gene movement that results in the erosion of colinearity between closely related species was recently investigated in a three-way comparison of the *Brachypodium*, rice, and sorghum genomes (Wicker et al. 2010). Gene capture by TEs, including *Helitrons*, was not found to have contributed significantly to gene movements within the grass family. On the other hand, TEs of many superfamilies, including *Helitrons*, were found at the borders of the noncolinear (i.e., mobilized) regions, suggesting that repair of TE-induced double strand breaks through synthesis-dependent strand annealing (SDSA) may have been involved in the change of position of genes in related genomes.

11.4 The Genetics of *Helitrons*

Being a member of the rare group of transposons that have been discovered computationally (Feschotte and Pritham 2007), it is not surprising that *Helitron* genetics trails its genomics. Yet, a genetic approach will be needed to identify a functional autonomous *Helitron* transposon, discern the actual mode(s) of transposition, assess the regulation of and by captured gene fragments, and elucidate other aspects of basic *Helitron* biology.

11.4.1 Transposition Mechanism: Rolling Circle and/or Cut-and-Paste?

A rolling circle replication mechanism has been proposed for the amplification of this novel class of transposons (Kapitonov and Jurka 2001). The putative autonomous *Helitrons* from the three genomes originally examined shared two conserved domains: the cross-kingdom DNA helicase domain and the replicator initiator proteins of RC plasmids and certain ssDNA viruses (Fig. 11.1a). Though still a hypothetical mechanism, RC replication is supported by the conserved structure of putative autonomous copies from several sequenced model plant genomes (Table 11.1).

The genome-wide distribution of *Helitron* elements favors a dispersive transposition model, although occasional *Helitron* clusters have been reported in some plant genomes (Lai et al. 2005; Yang and Bennetzen 2009a). Some peculiar head-to-head, head-to-tail, and tail-to-tail *Helitron* configurations have been identified in the maize genome (Du et al. 2008; Yang and Bennetzen 2009a), but they are composed of dissimilar *Helitrons* with similar terminal sequences, which differ

from the perfect head-to-tail *Helitron* configurations expected from a RC replication mechanism and, so far, found only in the *Myotis lucifugus* genome (Pritham and Feschotte 2007).

As discussed in Sect. 11.3.5, *Helitrons* have contributed to the frequent loss of genetic colinearity in related plant genomes. Many recently duplicated fragments in the grasses are bordered by transposable elements (TEs), including *Helitrons* (Wicker et al. 2010). Other chromosomal rearrangements, such as inversions, are also oftentimes associated with *Helitron* transposons. Of the 154 inversions identified between *Arabidopsis thaliana* and *Arabidopsis lyrata*, one-third are flanked by inverted repeats from *Helitron* elements (Hu et al. 2011).

In addition to RC replication, a *Helitron* cut-and-paste transposition mechanism, like the one used by most known DNA transposons, was recently proposed. Li and Dooner (2009) found that, unexpectedly, some maize *Helitrons* could excise somatically. The somatic excision products or footprints left by removal of a 6-kb *Helitron* consisted of a variable number of TA repeats at the prior insertion site, an unlikely consequence of a RC replication mechanism. Somatic excision products were also detected from other genic and agenic *Helitron* elements (Du et al. 2008; Li and Dooner 2009). This finding suggests that, like *Tn7* (Craig 2002) and *Mutator* (Walbot and Rudenko 2002), *Helitrons* may exhibit both replicative and excisive modes of transposition.

11.4.2 Gene Capture

Transduplication or the capture of host gene sequences, first reported for *Mutator* elements (Jiang et al. 2004; Talbert and Chandler 1988), is a common feature of several families of plant transposons (Dooner and Weil 2007). However, *Helitrons* may contribute the largest portion of transduplicated sequences in some plant genomes, like maize (Brunner et al. 2005b; Du et al. 2009; Lai et al. 2005; Morgante et al. 2005; Wicker et al. 2010; Yang and Bennetzen 2009a, b).

In contrast to the broad-spectrum of captured genes in maize, only a few genes have been captured by *Helitrons* in *A. thaliana* (Hollister and Gaut 2007; Yang and Bennetzen 2009b). Gene-capture by *Helitrons* is also a rare event in *Medicago*, *Brachypodium*, *sorghum*, and rice (Fan et al. 2008; Wicker et al. 2010; Yang and Bennetzen 2009b). No correlation has been found between the transcriptional orientation of the captured gene fragments and the orientation of the TE in which they are lodged. In fact, some *Helitrons* contain multiple genes with opposite transcriptional orientations (Lai et al. 2005; Lal et al. 2003; Wang and Dooner 2006; Wicker et al. 2010).

In spite of the well-documented transcriptional activities of genes captured by *Helitrons* from different plant species (Brunner et al. 2005b; Lai et al. 2005; Lal et al. 2003; Morgante et al. 2005 and see Sect. 11.4.3), no cases of functional full-length gene capture by *Helitron* elements have been reported. Although an almost

intact cytidine deaminase gene missing only the first six amino acids was found embedded in a maize *Helitron*, no transcripts corresponding to it were detected in any tissue examined (Xu and Messing 2006).

The capture of gene fragments from various genomic locations by the same *Helitron* may give rise to complex networks regulating the donor genes (Brunner et al. 2005b; Lai et al. 2005). The extent to which the host genome could benefit from these potentially deleterious effects (Du et al. 2009) is unclear.

11.4.3 Coevolution with the Host Genome

The potential role of *Helitrons* and other TEs in gene creation in plants has been recently reviewed by Dooner and Weil (2012).

Gene fragments captured by *Helitrons* originate from nonadjacent loci in the genome, yet they tend to be in the same transcriptional orientation relative to each other and to the *Helitron*'s *RepHel* gene. A large collection of gene-fragment-bearing *Helitron*s in maize show a notable bias in the orientation of gene fragments that is compatible with *Helitron* promoter-driven expression (Du et al. 2009; Yang and Bennetzen 2009a). Several chimeric transcripts containing exons from different genes ("exon shuffling") have been detected for maize *Helitrons* (Brunner et al. 2005b; Lai et al. 2005; Morgante et al. 2005). Though many of these transcripts contain premature stop codons in all reading frames and are unlikely to encode functional proteins immediately, *Helitrons* could have contributed to gene creation over evolutionary time (Brunner et al. 2005b). Expression of chimeric transcripts can also be driven by the promoter of the disrupted gene, rather than by a *Helitron* promoter. In maize, chimeric transcripts derived from genes captured by the inserted *Helitron* in the *sh2-7057* mutant are produced from the *sh2* promoter (Lal et al. 2003), rather than from a *Helitron* promoter.

The idea that TEs have been co-opted by the host as regulatory sequences has received considerable experimental support. Many *cis*-regulatory elements involved in transcriptional regulation have characteristics of TEs and some of them are *Helitrons*. For example, the *CArG* motif essential for the transcriptional activation of *LEAFY COTYLEDON2* (*LEC2*), a master regulator of seed development in *A. thaliana*, is located at the beginning of a *Helitron* element (*Helitron3*). This and other TE insertions located in the promoter region of *LEC2* were speculated to control the gene's specific expression pattern (Berger et al. 2011).

TE sequences are also found in transcripts, where they may play an unsuspected regulatory role. In *Arabidopsis thaliana*, more than 2,000 putative TE-gene chimeras, where a TE is found in at least one expressed exon, have been identified and compared to all TEs in a TE database (Lockton and Gaut 2009). *Helitron*-like sequences were strikingly underrepresented (2.4 %) in exons, contrasting with the high abundance (~20 %) of all other TEs. A similar pattern was found for the specific targets of the *MOM1* (*MORPHEUS' MOLECULE1*) regulator of transcriptional gene silencing in *Arabidopsis* (Numa et al. 2010). The majority of

MOM1 targets carry sequences related to TEs of both classes and are clustered at pericentromeric regions, suggesting that *MOM1* acts on regions of heterochromatin in the genome. *Helitron* remnants, on the other hand, were significantly underrepresented among *MOM1*-regulated transcripts. The authors suggested that, because *Helitrons* target active genes undergoing transcription, their low frequency among *MOM1*-target sequences may reflect exclusion of MOM1 from active chromatin environments. As major contributors to the evolution of plant genomes, more in-depth analyses are required to decipher the contributions of TEs to annotated protein-coding regions, an essentially unexplored field (Lal et al. 2009b).

11.4.4 Epigenetic Regulation

There is growing evidence that the proliferation of TEs in plants is under epigenetic regulation and that their biological properties are strongly affected by cycles of methylation and demethylation (Lisch 2009).

The past couple of years have seen a considerable increase in experimental data, mainly from *Arabidopsis*, on the methylation status of TEs. As shown in two earlier bisulfite sequencing studies (Gehring et al. 2006; He and Dooner 2009), *Helitrons* are heavily methylated at CG sites. In the first study, a *Helitron* inserted 4 kb upstream of the start site of the *Arabidopsis MEDEA* gene was heavily methylated, yet did not contribute to the allele-specific DNA hypomethylation in the endosperm (Gehring et al. 2006). In the second study, two maize *Helitrons* shown to be nonrecombinogenic despite the presence of multiple gene fragments were much more methylated than the adjacent recombinogenic gene-rich region (He and Dooner 2009).

Transcriptional reactivation of TEs in the mature pollen of *Arabidopsis* has been detected in microarray assays of TE expression profiles during development (Slotkin et al. 2009). In most tissues and stages, the ORFs of *Helitron2* and six other full-length TEs (including retrotransposons and DNA transposons) were either not expressed or expressed at a very low level, indicating that they are generally silenced. However, all seven full-length TEs examined were coordinately expressed in mature pollen. TE expression coincides with loss of DNA methylation and downregulation of the chromatin remodeler *DDM1*.

A recent study analyzed the contribution of TEs and small RNAs to gene expression variation in *A. thaliana* and *A. lyrata*, a closely related congener with a two to threefold higher copy number for every TE family examined, including *Helitrons* (Hollister et al. 2011). Reassessment of the TE content in the two species revealed that, unexpectedly, *Helitron*s were the highest copy number DNA transposons in both (Table 11.1). The 24-nt siRNA complements from the two species were compared in order to address the possible role of siRNA-guided transcriptional gene silencing in differential TE proliferation. *Helitrons* were found to be less often targeted by unique 24-nt siRNAs in *A. lyrata* than in *A. thaliana*, possibly explaining their higher copy number in the former. An almost concurrent reanalysis of DNA methylation, siRNA, and TE datasets from *Arabidopsis thaliana*

concluded that *Helitrons* actually contribute ~7 % of the annotated genome (Table 11.1) and, along with the *Tc1*/*mariner* superfamily, have the largest fraction (40–50 %) of unmethylated TE sequences (Ahmed et al. 2011).

Around a dozen *Arabidopsis* genes are imprinted, i.e., expressed in a parent-of-origin-dependent manner in the endosperm during seed development (Kermicle 1970). In a couple of cases, *Helitron* insertions have been implicated in imprinting. In a study on the association of TE methylation with gene imprinting during seed development in *A. thaliana*, TE fragments were found to be extensively demethylated in the endosperm (Gehring et al. 2009). Two imprinted members of the class IV homeodomain transcription factors contain remnants of *Helitron* elements at the 5′end. Although these genes showed reciprocal imprinting, i.e., predominant expression of the maternal allele in one and of the paternal allele in the other, methylation of the *Helitron* remnants was lost from the maternal alleles in both cases. Other imprinted genes are also neighbored by TEs. *AGL36*, a maternally expressed gene, contains remnants of *Helitrons* and other TE sequences within a 1.7-kb promoter fragment that is sufficient to confer parent-of-origin-specific expression of a reporter (Shirzadi et al. 2011). Paternally expressed genes, as well, are enriched for cis-proximal transposons, particularly for *Helitrons* (Wolff et al. 2011). It has been proposed that imprinting may have evolved from targeted methylation of TE insertions near genes followed by positive selection when the resulting expression change was advantageous (Gehring et al. 2009).

Whether a TE can exert a regulatory effect on a nearby gene obviously depends on the distance between the transposon and the gene. A methylated *AtREP2 Helitron* inserted 3.8 kb upstream of the imprinted *MEA* gene in the Col-0 and Ler-0 ecotypes of *Arabidopsis thaliana* was considered a candidate for imprinting control elements until ecotypes were found where *MEA* was still imprinted, though they lacked the upstream *Helitron* (Spillane et al. 2004). In a recent study relating gene expression to distance from the nearest TE in *A. thaliana*, average gene expression increased with distance up to about 2.5 kb (Hollister et al. 2011).

11.5 Perspective

The huge number of annotated *Helitron* transposons in plant genomes, including both putative autonomous elements and nonautonomous elements with and without gene fragments (Table 11.1), represents only the tip of the iceberg.

The molecular structure of the autonomous *Helitron* and the RC mechanism of transposition (Kapitonov and Jurka 2001) remain hypothetical, but are supported, respectively, by the conservation of structure of the putative autonomous element across evolutionarily widely divergent species and the identification of occasional head-to-tail configurations that make RC replication a credible transposition mechanism. Whether the RepHel protein is necessary and/or sufficient for RC transposition needs to be confirmed experimentally. The discovery of *Helitron* somatic excision products in maize (Li and Dooner 2009) suggests that *Helitrons* may transpose by both copy-and-paste and cut-and-paste mechanisms.

As is evident from successive sequence annotations of the same genome, determination of the overall *Helitron* contents in a given genome is a challenging and uncertain exercise (Feschotte and Pritham 2009). The conserved sequence and structure of the 3′ end of known *Helitrons* has served as the basis for the development of a number of ad hoc programs for specific genome-wide surveys of this highly divergent family of transposons. However, their cross-species applications are still not efficient in identifying *Helitrons* in new species and novel programs, possibly based on the recognition of conserved nucleotide patterns, are desirable for the efficient de novo identification of *Helitrons* from all genome sequencing projects.

Only a few cases of gene-fragment-bearing *Helitrons* have been identified in plants other than maize. The high frequency of gene fragment capture by maize *Helitrons* is enigmatic, but it has been suggested to result from a RepHel enzyme with a different replication/repair fidelity (Yang and Bennetzen 2009b). The identification and characterization of an autonomous *Helitron* in maize would be highly desirable because maize is an excellent experimental genetic system and has currently active elements, as is evident from several recently arisen mutations (Table 11.2).

The dynamic evolution of *Helitron* is best exemplified by the discovery in maize of a new group of *Helitron*-like sequences, designated *Heltir*, which end in perfect 37-bp TIRs (Du et al. 2009). The sequence variability of *Helitrons* and the presence in the genome of other forms, like *Heltirs*, complicate the accurate estimation of the contribution of this transposon superfamily to plant genomes.

References

Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H (2011) Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res 39:6919–6931

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M et al (2011) The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science 332:960–963

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Berger N, Dubreucq B, Roudier F, Dubos C, Lepiniec L (2011) Transcriptional regulation of Arabidopsis *LEAFY COTYLEDON2* involves RLE, a *cis*-element that regulates trimethylation of Histone H3 at Lysine-27. Plant Cell 23:4065–4078

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005a) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Brunner S, Pea G, Rafalski A (2005b) Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize. Plant J 43:799–810

Buti M, Giordani T, Cattonaro F, Cossu RM, Pistelli L et al (2011) Temporal dynamics in the evolution of the sunflower genome as revealed by sequencing and annotation of three large genomic regions. Theor Appl Genet 123:779–791

Cho SH, Addo-Quaye C, Coruh C, Arif MA, Ma Z et al (2008) *Physcomitrella patens DCL3* is required for 22–24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. PLoS Genet 4:e1000314

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

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J et al (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701

Chuck G, Meeley R, Irish E, Sakai H, Hake S (2007) The maize *tasselseed4* microRNA controls sex determination and meristem cell fate by targeting *Tasselseed6*/*indeterminate spikelet1*. Nat Genet 39:1517–1521

Cocca E, De Iorio S, Capriglione T (2011) Identification of a novel helitron transposon in the genome of Antarctic fish. Mol Phylogenet Evol 58:439–446

Craig NL (2002) *Tn7*. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington, D.C., pp 422–456

Craig NL, Craigie R, Gellert M, Lambowitz AM (2002) Mobile DNA II. ASM Press, Washington, D.C

Dooner HK, Weil CF (2007) Give-and-take: interactions between DNA transposons and their host plant genomes. Curr Opin Genet Dev 17:486–492

Dooner HK, Weil CF (2012) Transposons and gene creation. In: Fedoroff N (ed) Molecular genetics and epigenetics of plant transposons: sculpting genes and genomes. Wiley, Hoboken, NJ

Doutriaux MP, Couteau F, Bergounioux C, White C (1998) Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*. Mol Gen Genet 257:283–291

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

Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic *Helitron* landscape of the maize genome. Proc Natl Acad Sci USA 106:19916–19920

Du J, Grant D, Tian Z, Nelson RT, Zhu L et al (2010) SoyTEdb: a comprehensive database of transposable elements in the soybean genome. BMC Genomics 11:113

Fan C, Zhang Y, Yu Y, Rounsley S, Long M et al (2008) The subtelomere of *Oryza sativa* chromosome 3 short arm as a hot bed of new gene origination in rice. Mol Plant 1:839–850

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368

Feschotte C, Pritham EJ (2009) A cornucopia of Helitrons shapes the maize genome. Proc Natl Acad Sci USA 106:19747–19748

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329–341

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. Genome Biol Evol 1:205–220

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578

Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Baştürkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Peñalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438:1105–1115

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124:495–506

Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324:1447–1451

Gill N, SanMiguel P, Dhillon BDS, Abernathy B, Kim H et al (2010) Dynamic *Oryza* genomes: repetitive DNA sequences as genome modeling agents. Rice 3:251–269

Greco R, Ouwerkerk PB, Pereira A (2005) Suppression of an atypically spliced rice CACTA transposon transcript in transgenic plants. Genetics 169:2383–2387

Gui YJ, Zhou Y, Wang Y, Wang S, Wang SY, Hu Y, Bo SP, Chen H, Zhou CP, Ma NX, Zhang TZ, Fan LJ (2010) Insights into the bamboo genome: syntenic relationships to rice and sorghum. J Integr Plant Biol 52:1008–1015

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 pseudogenes. Plant Mol Biol 57:115–127

He L, Dooner HK (2009) Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for *Helitron* and retrotransposon insertions. Proc Natl Acad Sci USA 106:8410–8416

Holligan D, Zhang X, Jiang N, Pritham EJ, Wessler SR (2006) The transposable element landscape of the model legume *Lotus japonicus*. Genetics 174:2215–2228

Hollister JD, Gaut BS (2007) Population and evolutionary dynamics of *Helitron* transposable elements in *Arabidopsis thaliana*. Mol Biol Evol 24:2515–2524

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428

Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Proc Natl Acad Sci USA 108:2322–2327

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottilar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, Bergelson J, Carrington JC, Gaut BS, Schmutz J, Mayer KF, Van de Peer Y, Grigoriev IV, Nordborg M, Weigel D, Guo YL (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43:476–481

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan WZ, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275–1281

International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. Cell 91:501–510

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573

Kapitonov VV, Jurka J (1999) Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107:27–37

Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci USA 98:8714–8719

Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet 23:521–529

Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet 9:411–412, author reply 414

Kermicle JL (1970) Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics 66:69–85

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, Giroux MJ, Brendel V, Vallejos CE, Hannah LC (2003) The maize genome contains a *Helitron* insertion. Plant Cell 15:381–391

Lal SK, Georgelis N, Hannah LC (2009a) *Helitrons*: their impact on maize genome evolution and diversity. In: Bennetzen JL, Hake SC (eds) Handbook of maize: genetics and genome, vol 2. Springer, New York, pp 329–339

Lal SK, Oetjens M, Hannah LC (2009b) *Helitrons*: enigmatic abductors and mobilizers of host genome sequences. Plant Sci 176:181–186

Langdon T, Thomas A, Huang L, Farrar K, King J, Armstead I (2009) Fragments of the key flowering gene *GIGANTEA* are associated with helitron-type sequences in the Pooideae grass *Lolium perenne*. BMC Plant Biol 9:70

Le QH, Wright S, Yu Z, Bureau T (2000) Transposon diversity in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 97:7376–7381

Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12:615–627

Li Y, Dooner HK (2009) Excision of *Helitron* transposons in maize. Genetics 182:399–402

Lisch D (2009) Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol 60:43–66

Liu P, Sherman-Broyles S, Nasrallah ME, Nasrallah JB (2007) A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. Curr Biol 17:734–740

Lockton S, Gaut BS (2009) The contribution of transposable elements to expressed coding sequence in *Arabidopsis thaliana*. J Mol Evol 68:80–89

Lockton S, Gaut BS (2010) The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. BMC Evol Biol 10:10

McClintock B (1947) Cytogenetic studies of maize and Neurospora. Carnegie Inst Wash Yearbook 46:146–152

McClintock B (1952) Chromosome organization and gene expression. Cold Spring Harb Symp Quant Biol 16:13–47

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Pérez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452:991–996

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A et al (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y, Tanisaka T, Wessler SR (2006) Dramatic amplification of a rice transposable element during recent domestication. Proc Natl Acad Sci USA 103:17620–17625

Numa H, Kim JM, Matsui A, Kurihara Y, Morosawa T, Ishida J, Mochizuki Y, Kimura H, Shinozaki K, Toyoda T, Seki M, Yoshikawa M, Habu Y (2010) Transduction of RNA-directed DNA methylation signals to repressive histone marks in *Arabidopsis thaliana*. EMBO J 29:352–362

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman WD, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

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

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69

Scherrer B, Isidore E, Klein P, Kim JS, Bellec A, Chalhoub B, Keller B, Feuillet C (2005) Large intraspecific haplotype variability at the *Rph7* locus results from rapid and recent divergence in the barley genome. Plant Cell 17:361–374

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C,

Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Sherman-Broyles S, Boggs N, Farkas A, Liu P, Vrebalov J, Nasrallah ME, Nasrallah JB (2007) *S* locus genes and the evolution of self-fertility in *Arabidopsis thaliana*. Plant Cell 19:94–106

Shirzadi R, Andersen ED, Bjerkan KN, Gloeckle BM, Heese M, Ungru A, Winge P, Koncz C, Aalen RB, Schnittger A, Grini PE (2011) Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of *AGAMOUS-LIKE36*. PLoS Genet 7:e1001303

Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136:461–472

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

Spillane C, Baroux C, Escobar-Restrepo JM, Page DR, Laoueille S, Grossniklaus U (2004) Transposons and tandem repeats are not involved in the control of genomic imprinting at the MEDEA locus in Arabidopsis. Cold Spring Harb Symp Quant Biol 69:465–475

Surzycki SA, Belknap WR (1999) Characterization of repetitive DNA elements in Arabidopsis. J Mol Evol 48:684–691

Sweredoski M, DeRose-Wilson L, Gaut BS (2008) A comparative computational analysis of nonautonomous helitron elements between maize and rice. BMC Genomics 9:467

Talbert LE, Chandler VL (1988) Characterization of a highly conserved sequence related to mutator transposable elements in maize. Mol Biol Evol 5:519–529

Tempel S, Giraud M, Lavenier D, Lerman IC, Valin AS, Couée I, Amrani AE, Nicolas J (2006) Domain organization within repeated DNA sequences: application to the study of a family of transposable elements. Bioinformatics 22:1948–1954

Tempel S, Nicolas J, El Amrani A, Couee I (2007) Model-based identification of Helitrons results in a new classification of their families in *Arabidopsis thaliana*. Gene 403:18–28

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Tsukamoto T, Hauck NR, Tao R, Jiang N, Iezzoni AF (2010) Molecular and genetic analyses of four nonfunctional S haplotype variants derived from a common ancestral S haplotype identified in sour cherry (*Prunus cerasus* L.). Genetics 184:411–427

Turcotte K, Srinivasan S, Bureau T (2001) Survey of transposable elements from rice genomic sequences. Plant J 25:169–179

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Walbot V, Rudenko GN (2002) *MuDR*/*Mu* transposable elements of maize. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington, D.C., pp 533–564

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

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Wicker T, Narechania A, Sabot F, Stein J, Vu GT, Graner A, Ware D, Stein N (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Wicker T, Narechania A, Sabot F, Stein J, Vu GT et al (2008) Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. BMC Genomics 9:518

Wicker T, Buchmann JP, Keller B (2010) Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Res 20:1229–1237

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MT, Spillane C, Nordborg M, Rehmsmeier M, Köhler C (2011) High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. PLoS Genet 7:e1002126

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

Yang L, Bennetzen JL (2009a) Distribution, diversity, evolution, and survival of Helitrons in the maize genome. Proc Natl Acad Sci USA 106:19922–19927

Yang L, Bennetzen JL (2009b) Structure-based discovery and description of plant and animal *Helitrons*. Proc Natl Acad Sci USA 106:12832–12837