Plant MITEs: Useful Tools for Plant Genetics and Genomics

Ying Feng
Agriculture and Biotechnology College, Zhejiang University, Hangzhou 310029, China

MITEs (Miniature inverted-repeat transposable elements) are reminiscence of non-autonomous DNA (class II) elements, which are distinguished from other transposable elements by their small size, short terminal inverted repeats (TIRs), high copy numbers, genic preference, and DNA sequence identity among family members. Although MITEs were first discovered in plants and still actively reshaping genomes, they have been isolated from a wide range of eukaryotic organisms. MITEs can be divided into *Tourist*-like, *Stowaway*-like, and *pogo*-like groups, according to similarities of their TIRs and TSDs (target site duplications). In despite of several models to explain the origin and amplification of MITEs, their mechanisms of transposition and accumulation in eukaryotic genomes remain poorly understood owing to insufficient experimental data. The unique properties of MITEs have been exploited as useful genetic tools for plant genome analysis. Utilization of MITEs as effective and informative genomic markers and potential application of MITEs in plants systematic, phylogenetic, and genetic studies are discussed.

Key words: miniature inverted-repeat transposable element, MITE, transposable element, retrotransposon, evolution, plant

Transposable elements (TEs) are DNA segments that can insert into new chromosomal locations and often make duplicated copies of themselves in the process (*1*). Barbara McClintock discovered them in maize more than a half century ago as the genetic agents that are responsible for the sectors of altered pigmentation on mutant kernels (*2*). Such elements of various types have been found in many genomes, such as *Drosophila melanogaster*, yeast (*Saccharomyces cerevisiae*), *Escherichia coli*, *Caenorhabditis elegans*, and humans (reviewed in ref. *3*).

With the progress of large-scale DNA sequencing, it has become apparent that, far from being rare components of large genomes, TEs are the single largest component of the genetic material of most eukaryotes (*1*). They account for at least 45% of the human genome (*4*) and 50-90% of some grass genomes (*5*).

TEs in general are divided into two major classes according to their modes of transposition (*6*, *7*). Class I elements (retroelements), or retrotransposons, transpose by means of an RNA intermediate generated by reverse transcription (*8*, *9*). LTR retrotransposons are characterized by the presence of long terminal repeats. Non-LTR retrotransposons, such as LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements), lack these characteristic sequences (*8*, *9*). Class II elements transpose via a DNA intermediate and are characterized by the presence of terminal inverted repeats, or TIRs (*10*, *11*), such as the *Ac-Ds* family in maize and the *P* element in *Drosophila*. Several elements are difficult to classify, mainly because their mechanisms of transposition remain unknown (*12*). It is the case for miniature inverted-repeat transposable elements (MITEs), which were first described in plants (*13*) and later in other organisms (*14*).

Features of MITEs

MITEs were first described for grass genomes (*15*, *16*) and identified in several plant species, including maize (*15, 16, 17*), rice (*15, 16, 18, 19*), green pepper (*20*), and *Arabidopsis* (*10*, *11*), but have also been found in a wide range of other eukaryotic organisms, including *Caenorhabditis elegans* (*21*, *22*), fungi (*23*), mosquitoes (*12, 14, 24*), beetles (*25*), and certain vertebrates, like *Xenopus* (*26*), humans (*27*), and teleost fishes (*28*).

Structurally, MITEs are reminiscence of non-autonomous DNA (class II) elements with their small size (100-600 bp) and short TIRs (10-30 bp, the sequence between TIRs is A+T rich). However,

E-mail: qzhxue@zju.edu.cn

their high copy numbers (1,000 to 15,000 per haploid genome), target-site preference (for TA or TAA, usually 2-3 bp, but up to 9 bp according to the recent reports) (*29*, *30*), and the uniformity of related elements distinguish them from most previously described non-autonomous DNA elements (*1, 7, 31, 32, 33*).

Although MITEs as an outstanding class of repetitive sequences shares many common features, there are significant sequence similarities within MITE families but not between them (*33*). Members of different subfamilies within a family only resemble each other in structure but not sequence identity. Some MITE families, such as *Tourist*, *Stowaway*, *Bigfoot*, and *Micron*, are predicted to form distinct secondary structures (*15, 28, 30, 34, 35*) that are very useful for identifying these elements.

Another important feature of MITEs is their preference for inserting into low copy number sequences or genic regions, especially for the regions of 5′ and 3′ to a gene, where matrix attachment regions (MARs) are found (*11, 31, 33, 35, 36, 37, 38, 39, 40*). Tikhonov *et al.* (*37*) discovered 33 MITEs in a 225-Kb region of the maize genome flanking the gene *adh1*. None of the MITEs in this region was found within the 166-Kb (74%) of sequence that is occupied by LTR retrotransposons. In rice, of 18 *Kiddo* members in annotated sequences, 7 (40%) are found in introns, 6 (30%) reside less than 530 bp from a coding sequence (CDS), and 4 (20%) rest between 1 and 5 Kb from a CDS and no MITE seems to have hopped in coding sequences yet. The abundance of MITEs (and of DNA transposons in general) in the gene-rich regions of plant genomes was confirmed after the analysis of the complete genome sequence of *Arabidopsis* (which contains ~1,200 MITEs and ~1,000 other DNA transposons, together constituting ~6% of the genome) and partial genome sequences of rice (~100,000 MITEs and ~10,000 other DNA transposons, forming ~12% of the genome) (*12*,*41*). It has been noticed that in plants and mosquitoes, MITEs are frequently associated with wild-type alleles of genes, indicating a potential role of these elements in gene regulation and genome organization (*13, 14, 18*).

In addition, MITEs were also found, in several cases, to insert into each other. For instance, the first *Stowaway* element was found as an insertion in a sorghum *Tourist* element (*17*), whereas in another case a *Tourist* dimer was found in the same organism (*37*). In maize, some *Tourist* elements (name *Tourist*-Zm) were present as adjacent or nested insertion (*41*), and such MITE multimers were also reported in other organisms, including rice (*41*, *42*) and mosquito (*12*, *14*). In rice, among the 6,600 MITEs analyzed, more than 10% were present as multimers and some MITE families had a high frequency of self-insertions. For the *Castaway* and *Gaijin* families, this preference was caused by a high frequency of self-insertions; in contrast, *Ditto* elements were targeted by many other families of repetitive sequences; but the frequency of MITE insertions into class I or other class II elements was surprisingly low; and nested MITE multimers were arisen from independent insertion events (*41*). It has been believed that MITEs preferentially target other MITEs, rather than self-insertion (*12*).

To date, no MITEs have been found to encode any product required for their movement, and their mechanisms of transposition and accumulation in eukaryotic genomes remain poorly understood (*40*). Investigations of the ability of MITEs to transpose have been hindered by the fact that families with low sequence similarities may contain decayed MITE members that have lost their mobility. In contrast, highly homogeneous MITE families such as *Hbr* and *Kiddo* may still well be functional and may, therefore, be useful candidates for solving the mystery of MITE transposition (*39*). A common model for the transposition mechanism of *Tc1/mariner* elements has emerged from the functional study of a limited number of animal transposases (*43*). The N-terminal region of *Tc1/mariner* transposases contains DNA-binding domains that bind specifically to the TIRs (*43*). A C-terminal domain is characterized by an amino acid signature called the DDE/D motif that consists of two aspartic acid residues and a glutamic acid residue (or a third aspartic acid residue). This motif is required for catalysis of both the DNA cleavage and the strand transfer steps of the "cut and paste" transposition reaction (*43*).

Classification of MITEs in Plants

With the advantage of many genome sequencing projects, vast amounts of DNA sequence from a variety of plant and animal species, including rice, *Arabidopsis*, human, mouse, and *Drosophila melanogaster*, have become available for analysis. MITEs, with their high copy numbers, distinct structural features, and compact stature, are relatively easy to be mined from DNA se-

quence databases (*1, 44, 45*). And there are several new software tools, for instance, FindMITE (ref. *46*; http://www.biochem.vt.edu/aedes), Miropeats (*47*), REPuter (*48*), GENSOR (http://www.girinst.org/Censor_Server.html), RepeatMasker (http://repeatmasker.genome.washington.edu), and REON (ref. *49*; http://www.gentics.wust.edu/eddy/recon), to be used for mining the databases and rapidly identifying MITEs on the basis of their structural characteristics or sequence homology.

Table 1 Classification of Plant MITEs

Group type	Subclass	Species	Appprox. copy no.	TSD[a]	No. of TIRs	Approx. size (bp)	Reference(s)
Tourist-like	*Alien*	*Bell pepper& Solanaceae*	2,400	TWA	25	400	*20*
	At-mPIF2	*A.thaliana*	20	TWA	14	400	*44*
	ATTIR16T3A	*A.thaliana*	100	TWA	16	500	*50*
	ATTIRIX1	*A.thaliana*	70	TWA	16-40	350-400	*50*
	B2	*Z. Mays*	1,000	TWA	14	130	*15*, *41*
	Castaway	*O.sativa*	1,000	TWA	13	350	*18*
	Ditto	*O.sativa*	2,000	TWA	15	300	*18*
	Explorer	*O.sativa*	2,000	TWA	13	240	*18*
	Gaijin	*O.sativa*	3,000	WWW	17	180	*18*
	Hbr	*Z. Mays*	4,000	TWW	14	310	*31*, *34*
	Kiddo	*O.sativa*	ND[b]	TWA	14	270	*39*
	MathE1	*A.thaliana*	50	TWA	25	400	*51*
	MPing	*O.sativa*	14-70	TWA	37-268	430	*40*, *41*
	Os-mPIF2	*O.sativa*	150	WWW	14	260	*44*
	Various Tourist	*Grasses*	$> 5,000$	TWA	13-100	100-400	*15*, *16*
	Various Tourist Tourist	*A.thaliana*	> 300	TWA	> 14	300-500	*11*
	Wanderer	*O.sativa*	4,000	TWA	10	300	*18*
	Zm-mPIF	*Z. Mays*	6,000	TWA	13	350	*44*
Stowaway-like	*Stowaway*	Various flowering plants	ND	TA	> 10	70-350	*16*, *52*
	Various Stowaway	*O.sativa*	40,000	TA	20-150	100-350	*19*, *53*
	Various Stowaway	*A.thaliana*	300	TA	25-100	200-300	*11*
pogo-like	*AtATE*	A.thaliana	1,617	TA	23	742	*46*
	Emigrant	A.thaliana	500-1,000	TA	24	550	*10*
Unclassified	*Bigfoot*	*Medicago*	$> 1,400$	9 bp	9	136-319	*30*
	Crackle	*O.sativa*	500	8 bp	SIRs[c]	385	*35*
	Hairpin	*A.thaliana*	10	CTWAR	114	238	*54*
	Micron	*O.sativa*	100-200	TA	SIRs	400	*36*
	Pop	*O.sativa*	50	8 bp	SIRs	125	*35*
	Snabo-2	*O.sativa*	150	4 bp?	107	383	*52*
	Snap	*O.sativa*	100	7 bp	SIRs	170	*35*

[a] N= any nucletide; W = A or T; R= A or C;

[b] ND, not determined;

[c] These elements have no TIRs but have subterminal inverted repeats (SIRs).

To date, the identified MITEs can be further classified into many subclasses (Table 1). It is likely that additional MITE families remain to be discovered. In plants, elements of the *Stowaway* family have been identified in numerous dicots and monocots, whereas elements of other families have only been identified in narrow taxonomic groups (*16*). For instance, the *Tourist* element is known only from grasses (*15*, *17*), the *Emigrant* family only from *Arabidopsis* (*10*), and *Bigfoot* only from *Medicago* (ref. *30*; Table 1).

Two main obstacles were initially encountered in attempts to classify MITEs with respect to existing transposons. First, none of the available MITE sequences revealed clear-cut relationships with TEs that encode known transposases (autonomous elements). Second, and perhaps more significantly, few MITE family has been shown to be actively transposing (*1*, *32*). Owing to the fact that MITEs lack coding capacity and little sequence similarity exists between subfamilies, the classification of MITEs becomes complicated. Generally, in plants, the majority of characterized MITE families can be divided into two groups based on similarity of their TIRs and TSDs: *Tourist*-like and *Stowaway*-like (*33*, *36*). Much evidence links *Tourist* and *Stowaway* MITEs with two superfamilies of transposases, *PIF/Harbinger* and *Tc1/mariner*, respectively (*33*, *55*). However, a third archetypal element called *Emigrant* (*10*) is closely related to the *pogo*-like family of elements (*12*) and has different evolutionary history from the other two (*56*). Table 1 summarizes thorough collection of the previously published plant MITEs as well as those recently generated by the systematic mining of the complete genome sequences.

The discovery of *PIFa* in maize led to the recognition of a new superfamily of transposases, called *PIF/Harbinger* (*46*). These transposases are probably of ancient origin as they are distantly related to bacterial IS5 transposases, and have been identified in a wide range of eukaryotes, including various flowering plants, a fungus, and nematodes (*1*). Once *PIF*-like DNA elements were uncovered in the genomic sequences of rice, *Arabidopsis*, and nematodes, it became feasible to identify their associated *Tourist*-like MITEs by searching the respective genomic sequences for non-autonomous members (*1*).

Recently, the first active MITE was identified from rice through systematic computational database search and analysis of the mutability of a rice slender mutation of the glume (*33, 40, 57*). The miniature *Ping* (*mPing*) element is a sequence classified as a *Tourist*-like MITE of 430 bp, which is present in about 70 copies in *Nipponbare* and in about 14 copies in *93-11*. These all nearly identical *mPing* elements transpose actively in an *indica* cell-culture line (*33*) and are activated in cells derived from anther culture (*40*). An *mPing*-associated *Ping* element, which has a putative *PIF* family transposase, is implicated in the recent proliferation of this MITE family in a subspecies of rice (*40*).

Origin and Amplification of MITEs in Plants

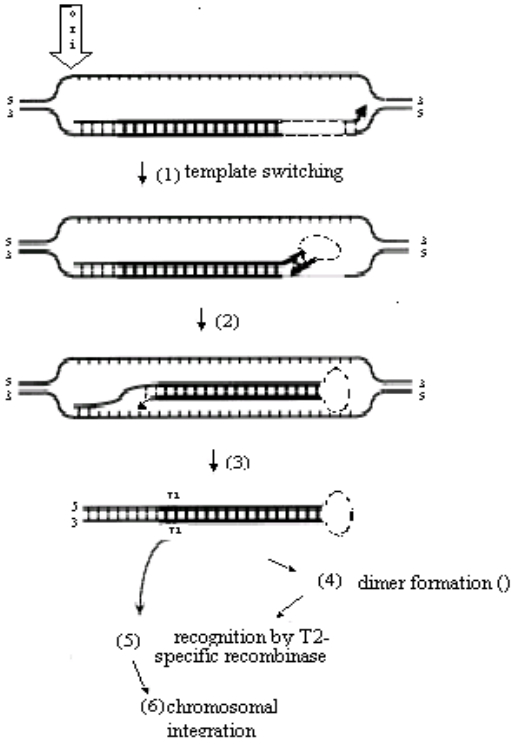
In the absence of coding sequences and an enzymatic activity, it has been difficult to determine how MITEs originate and how they attain their high copy numbers. Although descriptions of MITE families have proliferated in the literature, long-standing mechanistic issues concerning the birth, spread, transposition, and death of MITEs remain enigmatic.

A model suggested that the origin of MITEs involved aberrant DNA replication events when DNA polymerases encountered palindromic sequences as templates (*28*). In this model, the 3′ region of a nascent DNA strand may fold back, allowing DNA synthesis to reinitiate using the nascent DNA strand as templates. A stem loop byproduct (the *Angel* MITE) may result from this aberrant replication. After its excision from genomic DNA, the stem loop can then integrate into other genomic DNA locations with the help of recombinases, providing new sites for amplification of the MITE (Box1a). It was illustrated by the distribution of new rice-specific element, *Kiddo* members with respect to genic region (*39*). In this case, most *Kiddo* members clustered around the genic region and no MITE was in coding sequences, which indicate that the origin of the groups within the *Kiddo* family resides in introns. Such an arrangement could yield a very large number of *Kiddo* copies as a result of transcription and subsequent excision by splicing. Reverse-transcription of a proportion of the excised *Kiddo* RNA elements into DNA might confer transposition capability, permitting integration into nearby genomic DNA locations. Transposition into introns of other genes could lead to further propagation of *Kiddo* (*39*). But if the model is correct, excision of MITEs would not occur or remain as rare events, as previously suggested (*58*). Contradicting to this model, a more recent report has concluded that MITE did excise and leave footprints in the process (*29*). This

supports the notion that MITEs are DNA transposons and hence should be classified as class II elements.

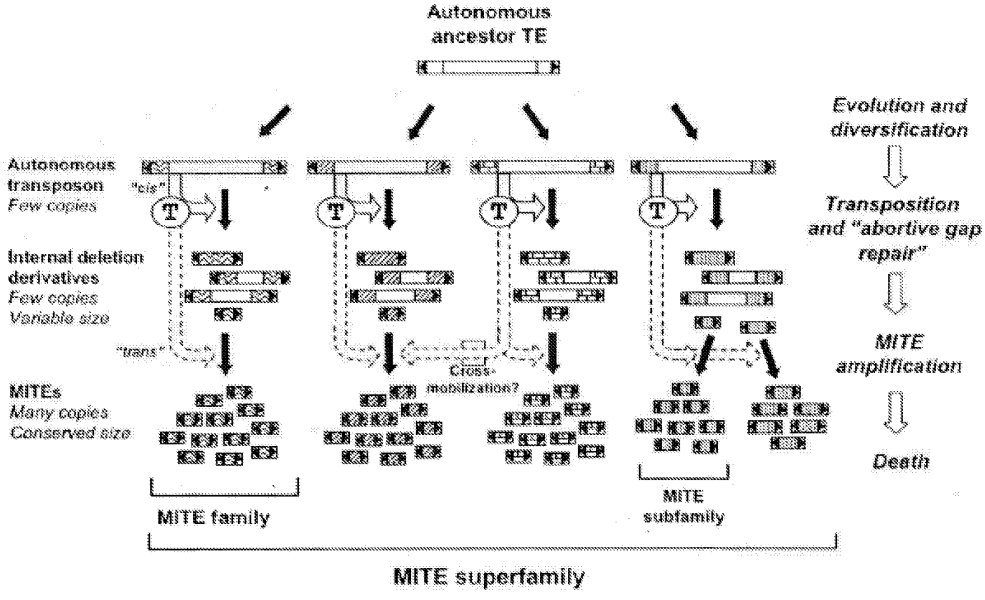
Box 1 Models for the origin and amplification of MITEs

a. A model for the formation and amplification of MITEs proposed by Izsvak *et al.* (*28*)



At the *top*, a replication bubble is shown, in which the DNA replication machinery passes through an *Angel* element (*solid line*) in the bottom parental DNA strand. The direction of DNA synthesis is from left to right. At step 1, the newly synthesized nascent strand may detach from the replication complex as a result of local destabilization due to the weak base-pairing at the A-rich central domain (*dotted line*) and the palindromic sequence of *Angel* that permits the formation of an intramolecular stem-loop, in which the A-rich sequences form the loop. At step 2, DNA synthesis reinitiates at the 38 end of the stem-loop, using the nascent strand as template. At step 3, DNA synthesis proceeds to form a duplicated *Angel* element in the form of an extrachromosomal stem-loop molecule that contains a palindromic copy of the sesquence between the replication origin and the loop of the *Angel* element. Intramolecular ligation and subsequent replication of the stem-loop DNA may result in dimerization of the element (step 4). Extrachromosomal *Angel* elements may be recognized by recombilase-like factors which may bind at the T2 motifs and facilitate chromosomal integration (steps 5 and 6).

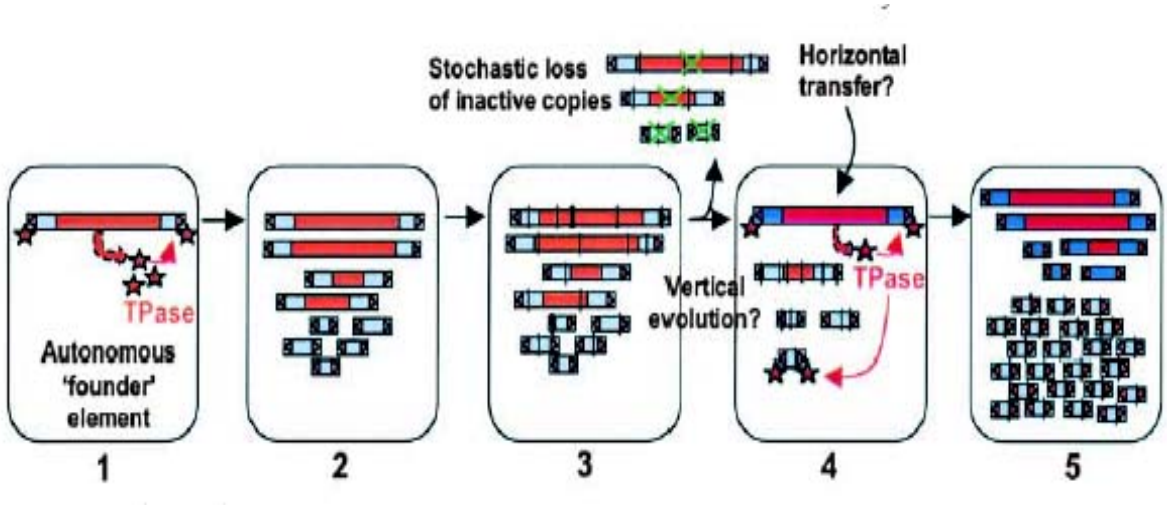
b. An alternative model proposed by Feschotte *et al.* (*1*, *32*)



The circled T stands for transposase. Transposase is known to mediate the formation of nonautonomous derivatives through mechanisms such as abortive gap repair (grey arrows). The subsequent amplification of one or few deletion

derivatives (*i.e.,* MITE amplification [dashed grey arrows]) is likely to be mediated by the ends of autonomous elements representing different subterminal sequences with identical or near-identical TIRs (black triangles).

c. A model for the amplification of *Stowaway* MITEs proposed by Feschotte *et al.* (*43*)



Step 1: an autonomous element transposes at relatively high frequency; Step 2: many newly internally deleted versions of the autonomous copy were produced because of frequent interruption and/or slippage during gap repair; Step 3: over time, both active and inactive copies are progressively degraded by point muations (vertical inactivation) and are stochastically lost or fossilized in the genome; Step 4: by chance, some of the decayed elements might preserve (or evolve *de novo*) sequences recognized by the transposase of a newly introduced autonomous element. The newly expressed transposase will thus be able to mobilize its own family members and distantly related MITEs; Step 5: sequence divergence between the MITE and its autonomous partner may favor the propagation of MITEs and allow their amplification to high copy numbers.

An alternative model was formulated that *Stowaway* elements originated by internal deletions from a larger autonomous element (like previously described nonautonomous DNA elements) and were amplified to high copy number by the transposase encoded by MITE (ref. *1, 32*; Box1b). There are three implications in the model. First, a MITE family is composed of subfamilies that have arisen from related autonomous elements in a single genome. A single type of autonomous element can give rise to one or multiple MITE families or can activate nonautonomous elements derived from a related autonomous element (*1*, *32*). The high sequence identity observed for many MITE families indicates that these families might have spread recently throughout their respective host genomes and the very high copy numbers attributed to many MITE families might be actually resulted from independent amplifications of different subfamilies in the same genome (*32*). This is illustrated best in rice (*40*), where *Stowaway* MITEs account for over 2% of genomic DNA (*38*). However, upon closer inspection it can be seen that there are more than 30 subfamilies of *Stowaway* MITEs and most of these has attained a copy number less than 1,000. Only a small number of these subfamilies have reached copy numbers that are significantly greater than 2,000 (*1*). A complex subfamily structure is also observed for *Tourist* MITEs (which make up ~3% of the rice genome). In contrast, larger genomes harbor very high copy numbers of MITE families. For example, there are over 6,000 copies of *mPIF* in maize that appear to have arisen from a single ancestral element (*1*). Accordingly, each *Stowaway* and *Tourist* MITE subfamily probably arose from the activity of related, but distinct, *mariner*-like and *PIF*-like autonomous elements, respectively (*1*, *33*).

Second, MITEs are seen as non-autonomous elements that originated from DNA transposons like previously descried nonautonomous elements in a two-step process. The first step, transposition of an au-

tonomous element, gives rise to various, internally deleted, non-autonomous derivatives. This step is very likely to be dependent on transposases and has been observed for other class II families, such as *Ac/Ds* or *P*-elements. In the second step, it is proposed that some derivatives can (for unknown reasons) amplify to high copy numbers. The hypothesis is supported by the recent discovery that rice and other plant genomes contain a tremendous diversity of *mariner*-like and *PIF*-like transposases (*12, 46, 55, 59*). Zhang *et al.*(*46*) isolated a maize *Tourist*-like MITE family called *miniature PIF* (*mPIF*) that shared several features with *PIF* elements, and suggested that *PIFa* and these *PIF*-like elements belong to a new eukaryotic DNA transposon superfamily that is distantly related to the bacterial IS5 group and are responsible for the origin and spread of *Tourist*-like MITEs (*59*). Phylogenetic analyses indicated that multiple divergent lineages of *mariner*-like MITE transposases can coexist within a single plant species and plant transposase sequences are monophyletic and extremely heterogeous (*55*).

The final implication of the model is the possible impact of MITE amplification on the evolution of autonomous elements. The birth and explosive amplification of MITEs could paradoxically be a death sentence for the transposase and consequently for the whole subfamily. However, selection would then lead to the diversification of the transposase favoring variants with altered binding sites, thus ushering in a new cycle of birth and death (*32*). With the identification of an active MITE family, such issues as the birth, spread, transposition, and death of MITEs can now be addressed experimentally. Some issues have been resolved already, for example, the prevalence of MITEs in single-copy regions primarily reflects targeting rather than selection (*33*).

To test the model, a semiautomated computational approach was used to identify and compare *mariner*-like element (MLEs) and the *Stowaway* MITEs in two draft genome sequences of rice, because several lines of evidence point to plant MLEs as the autonomous partners of the nonautonomous *Stowaway* MITEs (*43*). 34 different MLEs were found to group into three major clades and 25 families. More than 22,000 *Stowaway* MITEs were identified and classified into 36 families. On the basis of detailed sequence comparisons, MLEs were confirmed to be the best candidate autonomous elements for *Stowaway* MITEs. Surprisingly, however, sequence similarity between MLE and *Stowaway* families was restricted to TIRs and, in a few cases, to adjacent subterminal sequences. These data suggested that most of the *Stowaway* MITEs in rice were cross-mobilized by MLE transposases encoded by distantly related elements. The *Stowaway* MITEs were also gained supports from *mPing*, which was co-mobilized in cell culture with a closely, but not directly related, autonomous *Pong* element (*33*). In this model, the origin and amplification of MITEs were considered as two different steps that might be separated by a long period of time (ref. *43*; Box1c). The more time elapsed between these two steps, the more difficult it would be to recognize the coordination between a MITE family and an autonomous element. Based on recent research results, we may have to draw a conclusion that there are different mechanisms playing major roles in the origin and amplification of MITEs.

Application of MITEs in Plant Biological Researches

Effective and informative genetic markers

MITEs are useful in systematics and potentially useful as molecular markers because of their features, including high copy number, DNA sequence identity, polymorphism, and genic preference (*16, 18, 31, 34, 39*). The insertion sites of MITEs are frequently polymorphic with respect to their presence or absence at a particular locus between individuals of the same species. Because MITE excision seems to be extremely rare, MITE insertion polymorphisms have been successfully exploited as genetic markers (*31, 44, 60*). For example, *Hbr* has been successfully developed as a molecular marker and used for genetic characterization of a set of maize inbred lines (*31, 44, 60*). A new powerful marker technology called Inter-MITE polymorphism (IMP) was effectively applied in the fingerprinting of barley cultivars and for genetic similarity analysis (*44*). Many new MITE families, such as *Kiddo* that exists in a high copy number and share high sequence similarity in subclasses, also appear to have value as new molecular markers (*39*).

There are several merits of the MITE markers. First, compared with SSR, RFLP, and AFLP, MITE-marker possesses higher level of polymorphism (*60*). The level of MITE polymorphism as genetic markers is a reflection of their species specificity (*e.g.* the extent of restriction site polymorphism) and MITE

family specific factors. In the latter case, the extent of polymorphism reflects when each family spread through the population. That is, families that are still active or recently having been active will display higher levels of polymorphism than those active in the more distant past. However, because high sequence identity also correlates with recent amplification, it is anticipated that most families displaying high sequence identity will also be highly polymorphic in mapping population. Second, the ability to map markers with confidence is a function of the reproducibility of the protocol and the ability to unambiguously score segregating bands. In this regard, MITE markers are highly robust (*31*). Casa *et al.* (*31*) modified the AFLP procedure and developed transposon display (TD) technique to generate and display hundreds of genomic fragments anchored in *Hbr* elements. TD is a highly reproducible PCR-based protocol where multiple fragments are simultaneously detected using only a few primers. Third, compared with AFLP, MITE markers are more randomly distributed in the genome in macroscale (*31*). The distribution of markers in the genome has important implication for the general applicability and utility of the marker class. Randomly distributed markers are desirable as they provide maximum genome coverage. However, the distribution of MITE markers is consistent with the distributions of corresponding MITEs families, which usually prefer to the genic regions. Finally, the segregation distortion and non-parental bands of MITE markers are compatible with the RFLP markers and SSRs (*31*).

An additional source of variation for transposon-anchored markers is transposition. The newly discovered active MITE family *mPing* in rice (*33, 40, 57*) will play an important role as a marker. And if one of the parents harbors mobile elements, activity will be easy to discern by the appearance of an unusually high number of new bands or the loss of parental bands in the progeny.

But like the parent AFLP technique, MITE markers are mostly dominant, and their use will be limited in studies where discrimination of multiple alleles at a locus is required (*60*). And the ability to score MITE markers is determined, in part, by the number of amplified fragments displayed in each lane in an electrophoresis system. Adding selective bases to the restriction site-specific primer can reduce this in turn. In addition, the large number of products generated by MITE markers increases the probability of band overlapping among fragments that have the same size, *i.e.*, homoplasy (*60*).

Useful tools for molecular cloning

With the rapid progress of plant genome sequence project, more and more plant genome information becomes available. Insertional mutagenesis is the most suitable method for the systematic functional analysis of a large number of genes in the context of the whole plant. This system allows the production of many mutant lines at one time and the induced mutations can be easily detected by polymerase chain reaction (PCR). In *Arabidopsis* and rice, whose entire genomic sequences have been completed (*61, 62, 63*), several insertional mutagens were used to produce a large number of mutant lines. These include T-DNA, the maize transposable elements Ac/Ds and En/Spm, and plant retrotransposon, such as rice *Tos17* (*64*, *65*). The mutant populations induced by these mutagens are being used for molecular cloning, sometimes referred as forward and reverse genetics. For forward genetics, traditional transposon tagging is still an important method for cloning important genes for functional analysis.

Recently, the feasibility of tagging using *Tos17* has been demonstrated and several important genes have been cloned (*64*). *Tos17* becomes active under tissue culture conditions and has been used to develop a large-scale series of rice mutants in the rice genome project. Although this mutant series is regarded as a useful tool for analyzing gene functions, its gene-tagging efficiency is quite low (5-10%) owing to mutations induced by other factors under tissue-culture conditions. In addition, tissue-culture techniques for the *indica* subspecies have not been fully developed. Thus, the mobility of the *mPing* transposon in intact rice plants will provide a useful alternative tool for analyses based on reverse genetics in both the *indica* and *japonica* subspecies (*57*), and because *mPing* seems to be inserted in the genic region like other plant MITEs, it should be a useful molecular tool for gene isolation and gene knockout in these important crop plant species (*40*, *57*).

MITEs are known to be preferentially located in the close vicinity of genes. Active transposes under stress such as anther culture and gamma rays (*40*) will facilitate the tagging and thus isolation of functional genes. Although there has been no report on actual identification and isolation of particular genes, with the discoveries of more active plant MITEs, we are sure to see positive results in the near future.

MITEs and plant evolution

Currently, evolution scenarios of plant species are of great interests to biologists. It has been suggested that MITEs were important tools for evolution studies because their frequent occurrence in the regulatory regions of genes (*18*, *36*). For example, *Gaijin-So1* supplies almost the entire 3′-untranslated region of a sugarcane transporter cDNA; this element most likely supplies both the gene's polyadenylyation signal and site (*18*). MITEs with a known position in the genome may provide a phylogenetic signal for studying plant species evolution. If a number of taxa have a MITE inserted at exactly the same position on the chromosome, this is a good indication for common origins. The allelic diversity of recently activated or still active element in different species will reflect recent transposition events.

Kanazawa *et al.* (*53*) examined whether the presence or absence of MITEs may reflect evolutionary events such as speciation, expansion of habitats, and differentiation into ecotypes in wild rice species that share the same AA genome with cultivated rice. They have found that presence or absence of MITEs was highly conserved within each wild rice species except for *O. rufipogon*. In *O. rufipogon*, different patterns were detected in different ecotypes and the pattern was conserved within each ecotype. Conserved patterns were observed within each species even when different species showed overlapping distributions, such as *O. meridionalis* and *O. rufipogon* in Australia, and *O. barthii* and *O. longistaminata* in Africa.

Conclusion

Although MITEs were discovered just ten years ago, numerous classes of MITEs have been found in plants and other organisms. However, the identified active elements are extremely limited, such as the rice *mPing* element. Why there are so many elements discovered, but so little activity observed? What is the ultimate role of the MITEs to host genome and gene if it does exist? How do we learn about the evolution histories of the elements themselves? Peterson and Seberg (*29*) investigated the mode of *Stowaway* transposition and tried to trace the evolution of the element by the scrutiny of the phylogenetic tree. Turcotte and Bureau (*56*) suggested that the three main types of MITEs have different evolutionary histories despite their similarity in sequence structure. With the identification of more active elements and experimental studies, great progress is sure to be made on such issues.

Another challenge for the future is to determine whether, and if so how frequently, the sequence diversity created by MITE insertions has altered gene expression or gene products, as well as the fate of elements themselves after insertions. Although some studies have tried to touch these issues, the close proximity of *Kiddo* members to CDSs suggests that the insertion of these elements could probably modify transcriptional, splicing or translational regulation of the gene (*39*). The IR24 *rubq2* promoter that contains the *Kiddo* insertion has been shown to drive high levels of reporter gene expression in transient assays (*66*). Nevertheless, more systemic researches are definitely needed for the future in revealing the functions and functional consequences of MITEs as biological agents in extant host species.

References

1. Feschotte, C., *et al.* 2002. Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.* 3: 329-341.
2. McClintock, B. 1951. Chromosome organization and genic expression. *Cold Spring Harbor Symp. Quant. Biol.* 16: 13-47.
3. Berg, D. E. and Howe, M. H. 1989. *Mobile DNA*. American Society for Microbiology Press, Washington D.C., USA.
4. Lander, E. S., *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
5. Meyers, B. C., *et al.* 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research* 11: 1660-1676.
6. Finnegan, R. J. 1989. Eukaryotic transposable elements and genome evolution. *Trends in Genet.* 5: 103-107.
7. Kidwell, M. G. and Lisch, D. 1997. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. USA* 94: 7704-7711.
8. Kumar, A. and Bennetzen, J. L. 1999. Plant retrotransposons. *Annu. Rev. Genet.* 33: 479-532.
9. Bennetzen, J. L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* 42: 251-269.
10. Casacuberta, E., *et al.* 1998. Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. *Plant J.* 16: 79-85.
11. Le, Q. H., *et al.* 2000. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 97: 7376-7381.

12. Feschotte, C. and Mouchès, C. 2000. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Mol. Biol. Evo.* 17: 730-737.
13. Wessler, S. R., *et al.* 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Gen. Develop.* 5: 814-821.
14. Tu, Z. 1997. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* 94: 7475-7480.
15. Bureau, T. E. and Wessler, S. R. 1992. *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4: 1283-1294.
16. Bureau, T. E. and Wessler, S. R. 1994. *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6: 907-916.
17. Bureau, T. E. and Wessler, S. R. 1994. Mobile inverted-repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. *Proc. Natl. Acad. Sci. USA* 91: 1411-1415.
18. Bureau, T. E., *et al.* 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc. Natl. Acad. Sci. USA* 93: 8524-8529.
19. Turcotte, K., *et al.* 2001. Survey of transposable elements from rice genomic sequences. *Plant J.* 25: 169-179.
20. Pozueta-Romero, J., *et al.* 1996. Nonautonomous inverted repeat *Alien* transposable elements are associated with genes of both monocotyledonous and dicotyledonous plants. *Gene* 171: 147-153.
21. Oosumi, T., *et al.* 1995. Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 92: 8886-8890.
22. Surzycki, S. A. and Belknap, W. R. 2000. Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. *Proc. Natl. Acad. Sci. USA* 97: 245-249.
23. Yeadon, P. J. and Catcheside, D. E. A. 1995. Guest: a 98 bp inverted repeat transposable element in *Neurospora crassa*. *Mol. Gen. Genet.* 247: 105-109.
24. Tu, Z. 2000. Molecular and evolutionary analysis of two divergent subfamilies of a novel miniature inverted repeat transposable element in the yellow fever mosquito, *Aedes aegypti*. *Mol. Biol. Evo.* 17: 1313-1325.
25. Braquart, C., *et al.* 1999. DEC: a new miniature inverted-repeat transposable element from the genome of the beetle *Tenebrio molitor*. *Insect Mol. Biol.* 8: 571-574.
26. Unsal, K. and Morgan, G. T. 1995. A novel group of short interspersed repetitive elements (SINEs) in *Xenopus*: evidence of a specific target site for DNA-mediated transposition of inverted repeat SINEs. *J. Mol. Biol.* 248: 812-823.
27. Smit, A. F. A. and Riggs, D. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* 93: 1443-1448.
28. Izsvak, Z., *et al.* 1999. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *J. Mol. Evo.* 48: 13-21.
29. Peterson, G. and Seberg, O. 2000. Phylogenetic evidence for excision of *Stowaway* miniature inverted-repeat transposable elements in *Triticeae* (Poaceae). *Mol. Biol. Evo.* 17: 1589-1596.
30. Charrier, B., *et al.* 1999. *Bigfoot*: A new family of MITE elements characterized from the *Medicago genus*. *Plant J.* 18: 431-441.
31. Casa, A. M., *et al.* 2000. The MITE family heartbreaker (*Hbr*): Molecular markers in maize. *Proc. Natl. Acad. Sci. USA* 97: 10083-10089.
32. Feschotte, C., *et al.* 2002. Miniature Inverted-repeat Transposable Elements (MITEs) and their relationship with established DNA transposons. In *Mobile DNA II* (Craig, N., *et al.* eds.). American Society of Microbiology Press, Washington D.C., USA.
33. Jiang, N., *et al.* 2003. An active DNA transposon family in rice. *Nature* 421: 163-167.
34. Song, W. Y., *et al.* 1998. Identification and characterization of 14 transposon-like elements in the noncoding regions of members of the *Xa21* family of disease resistance genes in rice. *Mol. Gen. Genet.* 258: 449-456.
35. Akagi, H., *et al.* 2001. *Micron*, a microsatellite-targeting transposable element in the rice genome. *Mol. Gen. Genet.* 266: 471-480.
36. Zhang, Q., *et al.* 2000. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genie regions of maize. *Proc. Natl. Acad. Sci. USA* 97: 1160-1165.
37. Tikhonov, A. P., *et al.* 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* 96: 7409-7414.
38. Mao, L., *et al.* 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Research* 10: 982-990.
39. Yang, G., *et al.* 2001. *Kiddo*, a new transposable element family closely associated with rice genes. *Mol. Gen. Genet.* 266: 417-424.
40. Kikuchi, K., *et al.* 2003. The plant MITE *mPing* is mobilized in anther culture. *Nature* 421: 167-170.
41. Jiang, N. and Wessler, S. R. 2001. Insertion preference of maize and rice MITEs as revealed by the analysis of nested elements. *Plant Cell* 13: 2553-2564.

42. Tarchini, R., *et al.* 2000. The complete sequence of 340 kb of DNA around the rice *adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* 12: 381-391.
43. Feschotte, C., *et al.* 2003. Genome-wide analysis of *mariner*-like transposable elements reveals complex relationships with *Stowaway* MITEs. *Genetics* 143: 747-758.
44. Chang, R. Y., *et al.* 2001. Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. *Theor. Appl. Genet.*102: 773-781.
45. Amrani, A. E., *et al.* 2002. Genome-wide distribution and potential regulatory functions of *AtATE*, a novel family of miniature inverted-repeat transposable elements in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 267: 459-471.
46. Zhang, X., *et al.* 2001. *P*instability factor: an active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. *Proc. Natl. Acad. Sci. USA* 98: 12572-12577.
47. Parsons, J. D. 1995. *Miropeats*: graphical DNA sequence comparisons. *Computer Appl. Biosci.* 11: 615-619.
48. Kurtz, S. and Schleiermacher, C. 1999. *REPuter*: fast computation of maximal repeats in complete genomes. *Bioinformatics* 15: 426-427.
49. Bao, Z. and Eddy, S. R. 2002. Automated *De Novo* identification of repeat sequence families in sequenced genomes. *Genome Research* 12: 1263-1276.
50. Kapitonov, V. V. and Jurka, J. 1999. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* 107: 27-37.
51. Surzycki, S. A. and Belknap, W. R. 1999. Characterization of repetitive DNA elements in *Arabidopsis*. *J. of Mol. Evol.* 48: 684-691.
52. Chen, *et al.* 1997. Microcolinearity in *sb2*-homologous region of the maize, rice and sorghum genomes. *Proc. Natl. Acad. Sci. USA* 94: 3431-3455.
53. Kanazawa, A., *et al.* 2000. Inter- and intra-specific distribution of *Stowaway* transposable elements in AA-genome species of wild rice. *Theor. Appl. Genet.* 100: 327-335.
54. Adè, J. and Belzile, F. J. 1999. *Hairpin* elements, the first family of foldback transposons (FTs) in *Arabidopsis thaliana*. *Plant J.* 19: 591-597.
55. Feschotte, C. and Wessler, S. R. 2002. *Mariner*-like transposases are widespread and diverse in flowering plants. *Proc. Natl. Acad. Sci. USA* 99: 280-285.
56. Turcotte, K. and Bureau, T. 2002. Phylogenetic analysis reveals *Stowaway*-like elements may represent a fourth family of the IS630-Tc1-mariner superfamily. *Genome* 45: 82-90.
57. Nakazaki, T., *et al.* 2003. Mobilization of a transposon in the rice genome. *Nature* 421: 170-171.
58. Wessler, S. R. 1998. Transposable elements associated with normal plant genes. *Plant Physiol.* 103: 581-586.
59. Jurka, J. and Kapitonov, V. V. 2001. *PIFs* meets *Tourists* and *Harbingers*: a superfamily reunion. *Proc. Natl. Acad. Sci. USA* 98, 12315-12316.
60. Casa, A. M., *et al.* 2002. Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. *Theor. Appl. Genet.* 104: 104-110.
61. Bevan, M., *et al.* 2001. Sequence and analysis of the *Arabidopsis* genome. *Curr. Opin. Plant Biol.* 4: 105-110.
62. Stephen, A., *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. *ssp. japonica*). *Science* 296, 92-100.
63. Yu, J., *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. *ssp. indica*). *Science* 296: 79-92.
64. Hirochika, H. 2001. Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* 4: 118-122.
65. Kumar, A. and Hirochika, H. 2001. Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.* 6: 127-134.
66. Wang, J., *et al.* 2000. Structure, expression and promoter activity of two polyubiquitin genes from rice (*Oryza sativa* L.). *Plant Sci.* 156: 201-211.

Received: 21 April 2003
Accepted: 30 April 2003