Trans-activation and mobilization of a *Mutator*-like element in *Arabidopsis thaliana*

Yu Fu

Doctor of Philosophy

Department of genetics

School of life science

The Graduate University for Advanced Studies (SOKENDAI)

Department of Integrated Genetics National Institute of Genetics

2013

Contents

Summary

Transposable elements (TEs) are found in genomes of essentially every organisms examined. Mobile TEs are potentially mutagenic and deleterious for stability of the host genome, but most of TEs in eukaryotes are silenced by epigenetic mechanisms of host, such as RNA interference, histone modifications, and DNA methylation. The importance of DNA methylation in TE repression has recently been demonstrated in both mammals and plants by molecular genetic approaches. Less investigated research area is mechanisms of TEs to counter-act the host defense. Here I report activity of a plant TE to counter-act DNA methylation and silencing by the host.

VANDAL21 is a group of DNA TEs found in the genome of *Arabidopsis thaliana*. By structural similarity of encoded genes, *VANDAL21* has been classified as a member of *Mutator*-like elements (*MULE*s). Previous results of Southern analysis suggest that some of *VANDAL21* copies transpose in the background of reduced genomic DNA methylation in *A. thaliana* mutant *ddm1* (decrease in DNA methylation) (Tsukahara et al 2009). Using genomic DNA of the self-pollinated *ddm1* mutants, first I identified a mobile copy of *VANDAL21* family by PCR-based methods and whole-genome re-sequencing. I renamed the mobile *VANDAL21* copy *Hiun* (*Hi*; Japanese for “flying cloud”).

Unlike most of other mobile *MULE*s, the mobile copy does not have terminal inverted repeat (TIR). This is unusual because TIR has generally been thought to be the substrate for transposase. Despite the unorthodox structure, *Hi* excised and transposed as other typical *MULE*s with TIR. As other *MULE*s, *Hi* integrated with generating target site duplication around 9-bp long. In addition, it integrated preferentially near transcription start site. A unique behavior of *Hi* is that the integration has bias in the orientation, which may be related to the asymmetry in the terminal sequences of *Hi*. Excision of *Hi* often occurs with leaving the structure of the target locus before the integration, suggesting that the terminal positions in both sides are precisely determined even without TIR.

Hi has three open reading flames (ORFs), which I named *hiA*, *hiB* and *hiC*. The *hiA* encodes a protein with high sequence similarities to transposases found in other *MULE*s. Two other ORFs, *hiB* and *hiC*, do not have sequence similarity to any characterized proteins. These ORFs are silent in wild type background, where *Hi* is immobile. In order to see if transcriptional de-repression of these ORFs is sufficient for *Hi* mobilization, I introduced *Hi* transgene into wild type background. In the *Hi* transgene, all three ORFs are transcribed and the transgene induced excision of endogenous *Hi*. In addition, the *Hi* transgene induced loss of DNA methylation in terminal regions of endogenous *Hi*.

In order to know role of each ORF in the trans-acting effects of *Hi* transgene, I generated three types of transgenes with deletions for each of the three ORFs. The DNA de-methylation effect was abolished in transgene with deletion of *hiC* (Δ*hiC*), suggesting that *hiC* essential for the demethylation. On the other hand, Δ*hiB* (transgene with deletion of *hiB*) had the de-methylation activity indistinguishable from full length *Hi* transgene. Δ*hiA* (transgene with deletion of *hiA*) had de-methylation activity for one of the two terminals of endogenous *Hi*, which is upstream of *hiC*, but de-methylation activity was much reduced in the terminal region upstream of *hiA*.

I also examined effect of Δ*hiA* for the mobilization of endogenous *Hi*. In most of the transgenic lines, I could detect excision of endogenous *Hi*. The results were surprising, because *hiA* encodes a putative transposase, which presumably catalyses the transposition. However, subsequent analyses revealed that *hiA* transcript accumulates in Δ*hiA* lines, suggesting that endogenous *hiA* was de-repressed in the presence of Δ*hiA* transgene. Δ*hiA* transgene contains two ORFs, *hiB* and *hiC*. In order to test if *hiC* is sufficient for the transcriptional re-repression of endogenous *hiA*, I introduced a transgene with deletion of both *hiA* and *hiB* (Δ*hiA;B*). Δ*hiA;B* induced transcription of endogenous *hiA* and *hiB*. Δ*hiA;B* also induced excision of endogenous *Hi*, and DNA de-methylation of one terminal of endogenous *Hi*. These trans-acting effects are indistinguishable from those by Δ*hiA*, suggesting that *hiC*, rather than *hiB*, is responsible for these trans-acting effects of the *Hi* transgene.

In summary, I identified a mobile copy of *MULE*s without TIR and named that *Hiun* (*Hi*). When *Hi* is transformed into wild type plant, silent endogenous *Hi* copy was excised, suggesting that *Hi* is supplying factor(s) necessary for the transposition. *hiC*, one of the *Hi*-encoded gene, induced transcriptional activation, excision and DNA de-methylation of the repressed *Hi* copy. These trans-acting effects of *hiC* would contribute for counter-acting DNA methylation and silencing by the host.

Introduction

After identification of the first Transposable element (TE) in maize, many types of TEs have been identified in diverse organisms. However, only a few of those TEs have been shown to transpose. One of the causes is that hosts have mechanisms to repress TEs. DNA methylation is known as one of the TEs repression mechanisms. Changes in TEs activity are often correlated with the DNA methylation status; inactive TEs tend to be more methylated than active TEs (Chandler and Walbot 1986; Schläppi et al. 1996; Martienssen 1996). The importance of DNA methylation in TE control has also been demonstrated using mutants of *Arabidopsis thaliana*. In *A. thaliana* mutants with reduced genomic DNA methylation, a variety of silent TEs are de-repressed and transposed (Kato et al 2003; Lippman et al 2004; Tsukahara et al 2009; Mirouze et al 2009).

Intriguingly, some of TEs have mechanisms to counteract DNA methylation and silencing by the host. For example, *Suppressor-mutator* (*Spm*) element, a well-characterized TE in maize, encodes a tranposase gene *TnpA. TnpA* is known to induce loss of DNA methylation in regions controlling transcription of *Spm* (Schläppi et al 1994; Cui and Fedoroff 2002). That is thought to mediate the spontaneous activation of *Spm*, which is correlated with loss of DNA methylation.

Mutator element, another well-characterized TE in maize, also spontaneously changes its activity and DNA methylation in a coordinated manner (Martienssen and Baron 1994; Martienssen 1996). Like *Spm*, a silent *Mutator* element loses DNA methylation when active *Mutator* is present in the same genome (Brown and Suadaresan 1992). However, it is not understood how the interaction between active and inactive *Mutator* copies is mediated in trans.

MuDR, an autonomously mobile copy of maize *Mutator* family, contains two genes, *mudrA* and *mudrB*. The *mudrA* encodes the MURA protein, which is structurally similar to known trasnposases of other TEs (Eisen et al., 1994). In addition, *mudrA* is sufficient for excision of *Mutator*, further suggesting that

MURA functions as a transposase (Lisch et al., 1999). On the other hand, function of *mudrB* product is not well understood. TEs similar to *Mutator* are widespread in plant species and they are referred to as *Mutator*-like elements (*MULE*s) (Marquez and Pritham 2010). Open reading flames (ORFs) related to *mudrA* are generally found in mobile *MULE*s. Some of mobile *MULE*s also have additional ORF(s), such as *mudrB* in *MuDR,* but the structures of the proteins encoded in these ORFs are diverse and their functions remain largely unknown.

As is the case for most of class II (DNA type) TEs, *MuDR* element has relatively long terminal inverted repeat (TIR) of almost identical sequences. TIR is thought as important component for transposition of TEs. Interestingly, however, subgroups of *MULE*s without TIR have been found in the *A. thaliana* genome and they are classified as non-TIR *MULE*s (Le et al 2000; Yu et al 2000). Although the sequence analyses of *A. thaliana* genome suggest movement of these non-TIR *MULE*s in the past (Yu et al 2000), direct results describing their *de novo* transpositions are limited (Hoen et al 2006; Tsukahara et al 2009).

A group of non-TIR *MULE*, called *VANDAL21*, seems to transpose in background of reduced genomic DNA methylation (Tsukahara et al 2009). Here I identified an autonomous copy of *VANDAL21*, which is renamed *Hiun* (*Hi*). Even without TIR, *Hi* excised and transposed as canonical mobile *MULE*s with TIR. Interestingly, *Hi* transgene induced loss of DNA methylation, transcriptional activation, and excision of the endogenous *Hi* copy. Most importantly, these trans-acting effects of *Hi* do not depend on putative transposase, but depend on another function unknown gene encoded in *Hi*. These trans-acting effects would be very efficient for counteracting DNA methylation and silencing by the host.

Results

Identification of mobile *VANDAL21* copies

Database search suggested there are seven copies of *VANDAL21* elements with very similar sequences in the genome of *A. thaliana*. Consistent with that, Southern analysis revealed seven bands for that group of *VANDAL21* (Tsukahara et al 2009). At previous report, it was shown that the band pattern changes in *A. thaliana* plants self-pollinated multiple times in *ddm1* (*decrease in DNA methylation 1*) mutant backgrounds (Tsukahara et al 2009), suggesting mobility of one or more copies of the *VANDAL21* members (Tsukahara et al 2009). *A. thaliana ddm1* mutation generally induces loss of DNA methylation in TEs, which causes mobilization of diverse TEs (Lippman et al 2004; Tsukahara et al 2009; Mirouze et al 2009). In order to know which of *VANDAL21* copies are mobile, two methods were used: suppression PCR and whole genome re-sequencing (see Methods for details).

In total, I identified 53 *de novo* insertions of *VANDAL21*s (38 by genome re-sequencing and 15 by suppression PCR) in the self-pollinated *ddm1* lines (Table1). Of these 53 insertions, 50 correspond to one copy (*AT2TE42810*; Fig1 & Table1) of *VANDAL21* element. The remaining three insertions correspond to another copy (*AT4TE15615*; Table1). For the other five copies of *VANDAL21*, no new insertion has been identified. In the following parts, I concentrate on the most active copy, *AT2TE42810*, which is renamed *Hiun* (*Hi*, Japanese for “a flying cloud”).

Structure of *Hi*

Hi is 8177bp long, and includes three ORFs, *At2g23500*, *At2g23490*, and *At2g23480* (Fig 2). One ORF (*At2g23500*; *hiA*) encodes a gene with high sequence similarities with MURA-type transposases, which are generally found in *MULE*s. Two other ORFs (*At2g23490*; *hiB* and *At2g23480*; *hiC*) do not have

sequence similarity to any characterized genes. An unorthodox feature is that, unlike other typical mobile *MULE*s (Singer et al 2001; Chalvet et al 2003; Xu et al 2004; Gao 2012), TIR of this TE is extensively degenerated (Fig 3; Fig 4), showing the characteristics of non-TIR *MULE* (Yu et al 2000). Despite the unorthodox structure, *Hi* transposed in *ddm1*. For studying the transposition manner of this TE, *ddm1* mutant was used.

Integration and excision of *Hi* at *ddm1* mutant

Generally, *Mutator* elements preferentially transpose into 5' region of a gene (Hardeman and Chandler 1989; Dietrich et al 2002; Liu et al 2009). That was also the case for *Hi*; most of the integration sites are localized around transcription start sites of genes (Fig 5). Interestingly, *Hi* showed a biased integration in the orientation (Fig 5; Table 1). Such bias in the orientation has not been reported in the *Mutator* element (Brown et al 1989). The bias in the orientation of *Hi* integration might be related to the diversity in the sequences of the two terminal regions.

Typical *MULE*s generate 9bp Target Site Duplication (TSD) at integration site (Lisch 2002). Therefore I examined whether *Hi* generates TSD or not. I identified the flanking sequences of randomly chosen 9 integration sites of *Hi*. Seven of them generated 9 bp TSD. And remaining 2 integration sites have 2 bp and 10 bp TSD, respectively (Fig 6).

MULEs belong to DNA type TEs, which normally have TIR. Because *Hi* does not have TIR, an interesting question would be whether *Hi* excision could be detected or not. I then examined the excision by PCR using primers for both of the flaking regions of the original *Hi* locus (Fig 7). By this assay, I could detect *Hi* excision in all independent *ddm1* lines examined (Fig 7). On the other hand, the excision could not be detected in any of *DDM1* sibling lines, confirming that *ddm1* mutation induces the *Hi* excision.

I further confirmed the excision by sequencing the PCR products (Fig 8).

Interestingly, even though *Hi* does not have TIR, many of the excision products showed excision around the terminal sites predicted from the integrated copies. Even TSDs are lost in significant part of the excision product. In summary, these observations suggest that the TIR is dispensable not only for integration, but also for the precise excision of this element in the defined termini. The transposition of *Hi* occurred in the manners comparable to those of typical DNA type TEs with TIR.

Trans-activation of repressed endogenous *Hi* by transgene

Hi transposed and excised in *ddm1*. For that, *Hi* is thought as an autonomous TE copy. It is known that an active autonomous TE copy can activate other repressed copies to transpose, if they co-located at same host genome. In order to test whether *Hi* is an autonomous copy or not, I introduced cloned *Hi* copy into wild type plant to check whether *Hi* transgene could induce excision of the endogenous *Hi* copy. For distinguishing the transcripts of the transgene from the endogenous *Hi* copy, I introduced silent mutations into the coding regions of the 3 ORFs in transgene (Fig 9). In all examined transgenic lines, excision of the endogenous copy was detected. (Fig 10). Neither transgenic line of empty vector nor wild type line showed excision of the *Hi*, confirming that *Hi* transgene triggered mobilization of the endogenous copy. It suggests *Hi* might be an autonomous copy. I also examined *Hi* activity in self-pollinated progeny of one of the *Hi* transgenic lines. I could detect excision of the endogenous *Hi* in the almost all progenies with the transgene. On the other hand, I could not detect the excision in progeny without the transgene, also confirming that the transgene is responsible for keeping the excision (Fig 11).

Interestingly, in the presence of the transgene, DNA methylation level was reduced in both termini of the endogenous *Hi* (Fig 12; Fig 13). The loss of methylation is more extensive in non-CpG sites than in CpG sites. When the transgene was segregated apart in the self-pollinated progeny of the transgenic line, the terminal regions were remethylated (Fig 13), which is associated with its immobilization (Fig 11).

Trans-activation of repressed endogenous *Hi* by *hiC*

The results shown above demonstrate that *Hi* transgene induced excision of the endogenous copy in trans, which is associated with loss of DNA methylation in the terminal regions. In order to further dissect the role for each of the ORFs in *Hi*, I generated transgenes with deletion in each ORF (Fig 9).

Surprisingly, Δ *hiA*, a transgene with deletion of putative transposase (*hiA*), induced excision of the endogenous *Hi* (Fig 14). However, the data of RT-PCR showed that even at the transgenic lines of Δ *hiA*, *hiA* transcript was detected (Fig 15). It suggests excision of the endogenous *Hi* copy seems occurring by *hiA* originating from the endogenous *Hi*.

The data of RT-PCR and qRT-PCR indicated that even in the transgenic lines of Δ *hiA* and Δ *hiB,* the transcripts of *hiA* and *hiB* were detected, respectively (Fig 15; Fig 16; Fig 17). By contrast, in the transgenic lines of Δ *hiC, hiC* transcript was undetectable (Fig 16; Fig 18).

I then examined effect of *hiC* at DNA de-methylation of the terminal sequence of the endogenous *Hi*. In the transgenic lines of Δ *hiC*, reduction of DNA methylation did not occur in either of the terminal regions (Fig 12), suggesting that *hiC* is essential for DNA de-methylation at the endogenous *Hi* copy. By contrast, in the transgenic lines of Δ *hiA* and Δ *hiB* , DNA de-methylation of the 3' region occurred at same level to *Hi*-TG, the transgene with full-length of endogenous *Hi* (Fig 12). It suggests *hiA* and *hiB* are dispensable for DNA de-methylation at the 3' region. DNA de-methylation of the 5' region seems more complex. A partly DNA de-methylation occurred in the transgenic lines of Δ *hiA* (Fig 12), suggesting *hiA* is necessary for DNA de-methylation of the 5' region.

Interestingly, although *hiC* is essential for activation of transcript and DNA de-methylation, it is not essential for excision of the endogenous copy. In the

transgenic lines of Δ*hiC,* excision was detected in some of the transgenic lines (Fig 14).

For confirming the effects of *hiC*, Δ*hiA;B,* the transgene with deletion of *hiA* and *hiB* was used (Fig 9). First, I detected excision of the endogenous copy (Fig 14). And the transcripts of *hiA* and *hiB* were detected at the transgenic lines (Fig 16; Fig 19). DNA de-methylation of the endogenous *Hi* was also detected (Fig 12). It suggests *hiC* could induce DNA de-methylation, transcription, and excision of the endogenous *Hi* copy.

Discussion

Trans-activation of *Hi* by *hiC*

Here I report identification of a mobile copy of *Mutator*-like element (*MULE*) without TIR. When this copy, named *Hiun* (*Hi*), is transformed into WT, silent endogenous *Hi* copy was excised, suggesting that *Hi* is the autonomously mobile copy.

Among three ORFs *Hi* contains, *hiA* encodes a protein similar to transposases in other *MULE*. Most importantly, *hiC*, another ORF of *Hi*, plays roles for transcriptional activation of other ORFs in *Hi*, and mobilization of *Hi*. These trans-acting effects would be very efficient for counter-acting DNA methylation and silencing by the host.

Molecular function of *hiC*

I have shown that *hiC* induced excision, transcriptional activation and DNA de-methylation of endogenous *Hi* copy. But the molecular mechanism remains to be clarified. One thing to clarify would be whether the *hiC* function is mediated by the encoded protein. Genome-wide analyses for effects of *hiC* on transcription and DNA methylation, would also be informative.

Reactivation of endogenous *Hi*

Hi TG induced reactivation of repressed endogenous *Hi*. Interestingly, however, endogenous *Hi* activated by the *Hi* TG was re-methylated and silenced after removal of the transgene (Fig 11; Fig 13; Fig 20). Why the reactivated endogenous copy could not keep the activity? One possibility is that *Hi* TG could not reactivate the endogenous copy completely. Although methylation at non-CpG sites was almost completely lost, methylation of the CpG sites only reduced slightly (Fig 12; Fig 13). The methylation at the CpG sites might prevent complete reactivation of the endogenous copy.

Non-autonomous copy of *Hi*

In this study, I also studied the transposition of other 6 copies of *VANDAL21*, which have sequence similarity to *Hi*, in *ddm1*. I could only find one copy transposed (*AT4TE15615*; Table1). Because the coding regions have many end codes that might prevent intact translation, the copy is thought as a non-autonomous copy of *Hi*. It is still not clear why transpositions of other 5 copies were not detected in *ddm1*.

Mode of transposition of *Hi*

Interestingly, even without TIR, most of the excision occurred in a manner to keep the original sequence before integration (Fig 8). That was found for both *Hi* mobilized in *ddm1* mutation and in the transgenic lines of *Hi* TG. That mode of excision, recovery of the target site to the original sequence, would be less deleterious to the host, and could be advantageous for survival of the element itself. *Hi* includes three ORFs, one of which encodes a protein similar to MURA transposase found in other *MULE*s. As MURA generally recognizes terminal sequences (Benito and Walbot 1997). It is interesting if one transposase recognizes two different sequences in the two termini and catalyzes precise excision of the defined termini.

Integration of *Hi* at transgenic plants

Though the excision of endogenous *Hi* was found at transgenic lines of *Hi* TG (Fig 10), I could not detect *de novo* integration of *Hi* at the lines by Southern analyses (Fig 21). Southern analyses is less sensitive than PCR or whole-genome re-sequencing. In addition, frequency of excision can also be lower in the transgenic lines than in the *ddm1* mutant. De novo integration in the transgenic line might be detected by whole genome re-sequencing of the transgenic lines.

Trans-acting controls and cis-acting conditions

Deletion of *hiA* in the *Hi* transgene affected efficiency of DNA de-methylation for the upstream region of *hiA* in the endogenous copy. The effect may have mechanistic link to effect of active *MuDR*. Active *MuDR* induces loss of methylation in one of the terminal regions, which is upstream of *mudrA* (Lisch et al 1999).

My results suggest involvement of *hiC* on transcriptional de-repression of *hiA* and excision of *Hi*. Interestingly, however, although the *Hi* transgene was transcribed and induced excision of endogenous *Hi*, I could not detect any excision activity of the transgenic *Hi* copy by PCR at some transgenic lines, even in the conditions where endogenous *Hi* was excised (Fig 22). The sequences are identical between these two *Hi* copies. Trans-acting factors and TE sequence do not seem to be the only determinants of its mobility; other cis-acting conditions, such as chromatin states or locus-specific genomic environment may also be involved.

Figures

De novo Integration sites of *Hi*

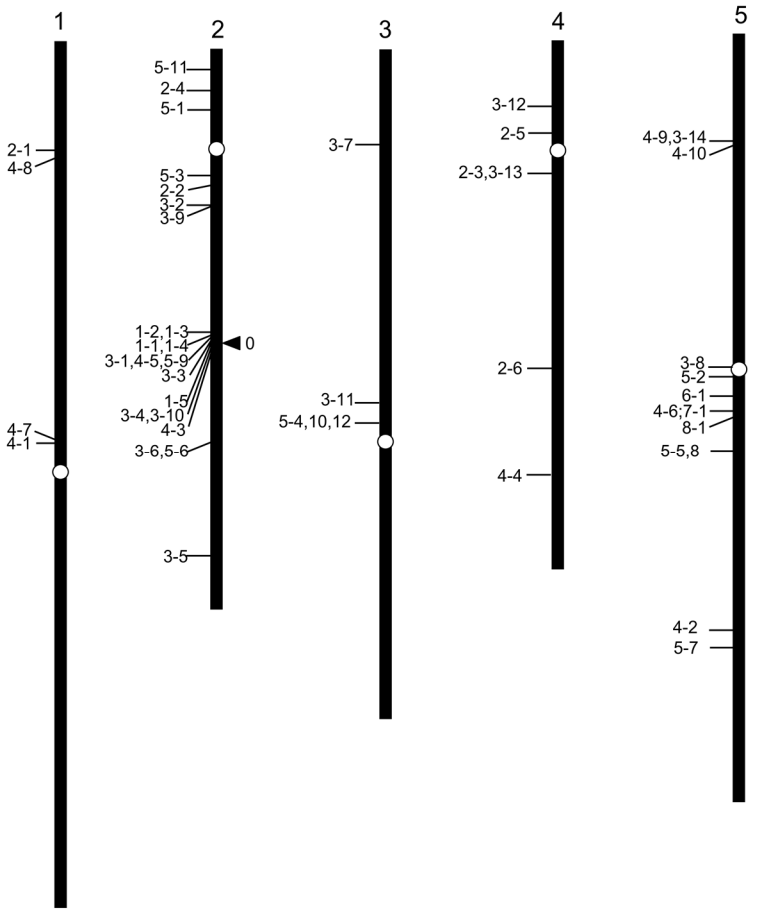


Figure 1

De novo integration sites of mobile *Hi* within the *A. thaliana* genome. Black bars represent 5 chromosomes of *A. thaliana.* 0, written at right side of chromosomes with black arrowheads, represent original loci of *Hi*. Loci for *de novo* integrations are shown in the left side of chromosomes. The first Arabic numerals (from 1 to 8) reflect different *ddm1* lines examined

Schematic diagram for structures of *Hi*

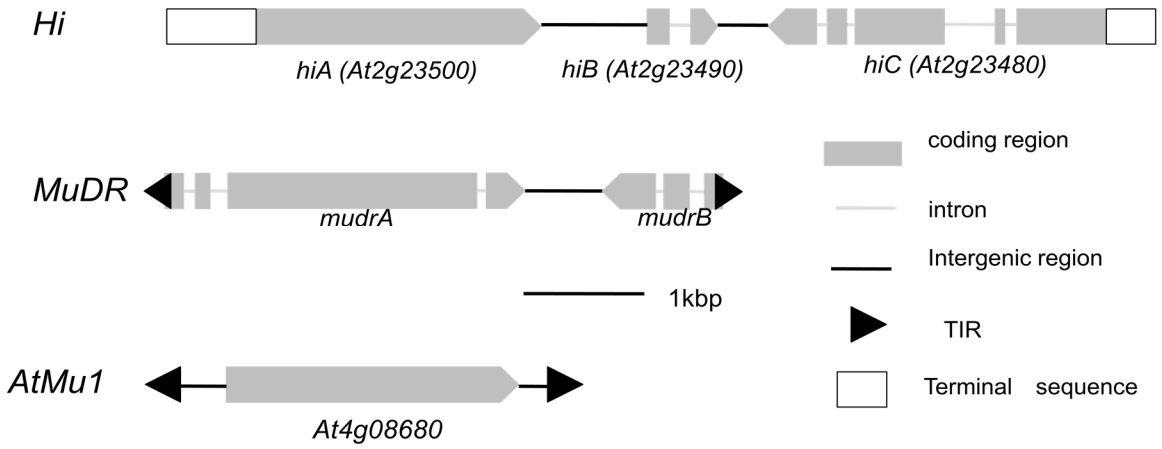


Figure 2

Schematic diagram for structures of *Hi* and two other mobile *MULE*s, *MuDR* and *AtMu1*. *MuDR* is the most well-characterized mobile *Mutator* in maize. *AtMu1* is an *A. thaliana MULE*, which is also shown to be mobile (Singer et al 2001; Tsukahara et al 2009). *Hi* encodes three ORFs. One of them, *hiA*, has sequence similarity to *mudrA*, a transposase encoded in *MuDR*. Functions are unknown for proteins encoded in the remaining two ORFs, *hiB* and *hiC*.

Dot plot of *Hi* and *AtMu1*

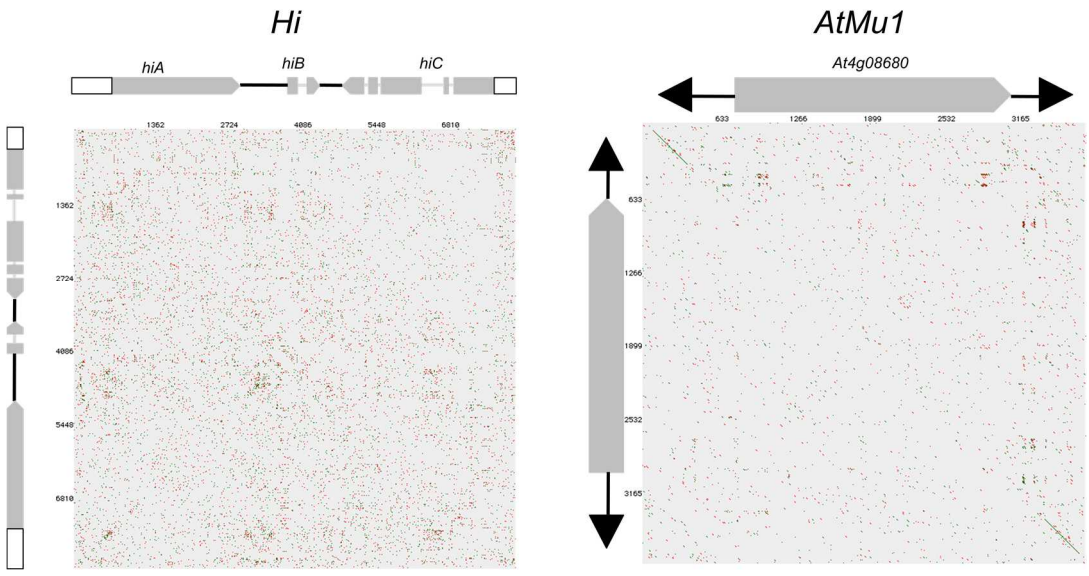


Figure 3

A dot plot analysis of *Hi* and *AtMu1*. The full length sequence of *Hi* and *AtMu1* were compared to its complement sequence. And *Hi* does not have detectable TIR (Terminal Inverted Repeat). (Window size: 10bp, mismatch limit: 1bp)

Terminal and flanking sequence of *Hi*

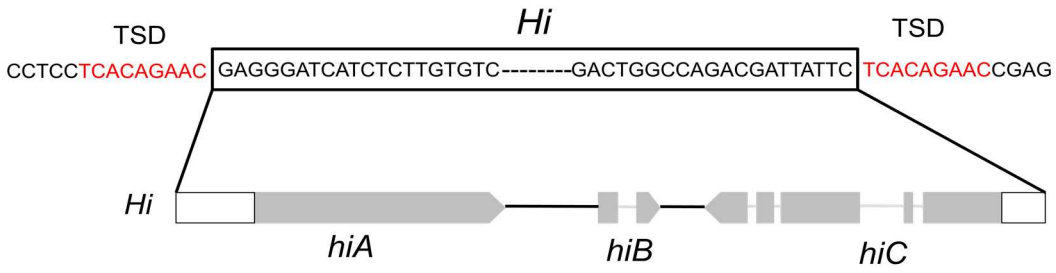


Figure 4

Terminal regions of *Hi* and flanking sequence with 9-bp of target site duplication (TSD). The 9-bp TSD is also found in most of neo-insertion sites (Fig 6). Any inverted repeats (>5bp) can find at outmost terminal sequences of *Hi*.

De novo integration sites of *Hi* in relation to flaking transcription units

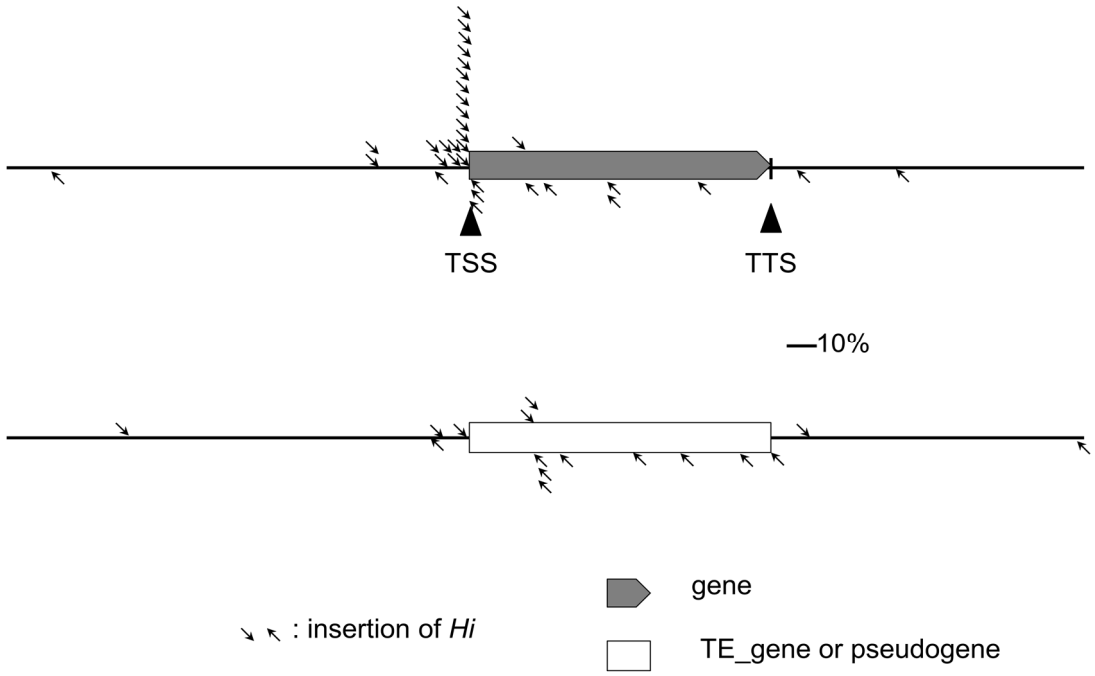


Figure 5

De novo integration sites of *Hi* in relation to flaking transcription units. Positions of integrations are normalized by length of the transcription units. Rightward and leftward arrows indicate insertions with 5' to 3' and 3' to 5' orientations of *Hi,* respectively. Insertions flanking pseudogenes and TE genes are shown in the bottom, and those flanking canonical genes in the top.

Target Site Duplication of *Hi*

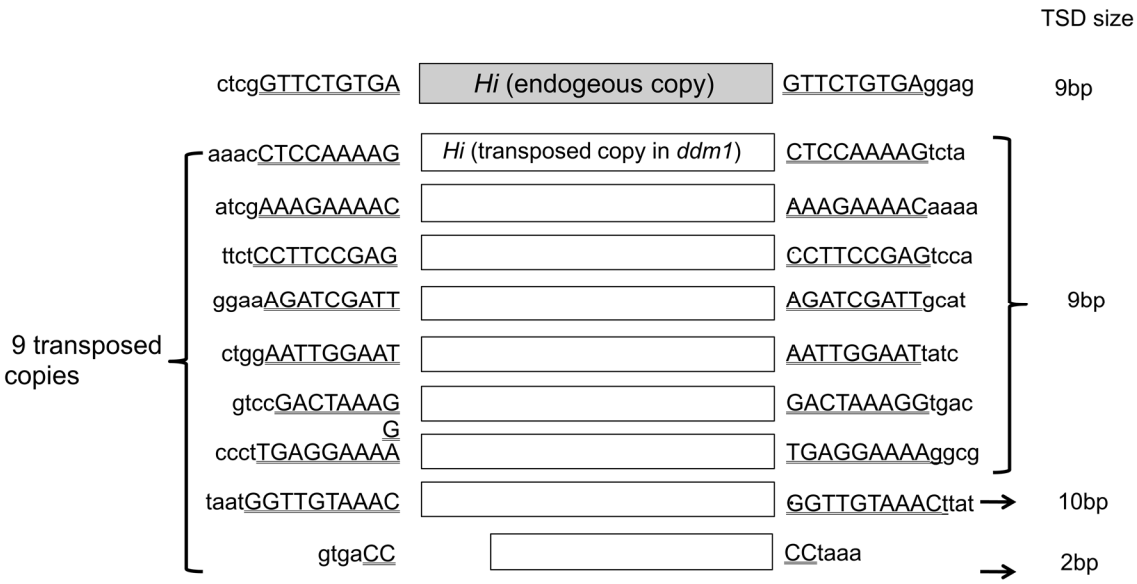


Figure 6

The flanking sequence of endogenous *Hi* and 9 *de novo* integration sites in *ddm1*. Seven integration sites have 9-bp perfect TSD. Remaining 2 integration sites have 2-bp or 10-bp perfect TSD, respectively.

Excision of the endogenous *Hi* copy in *ddm1*

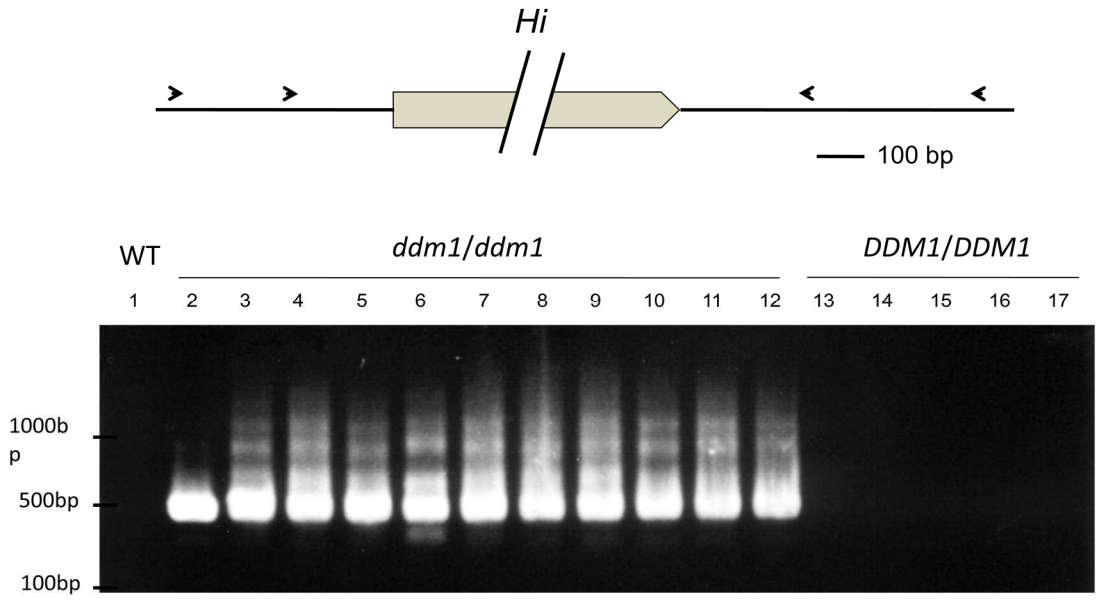
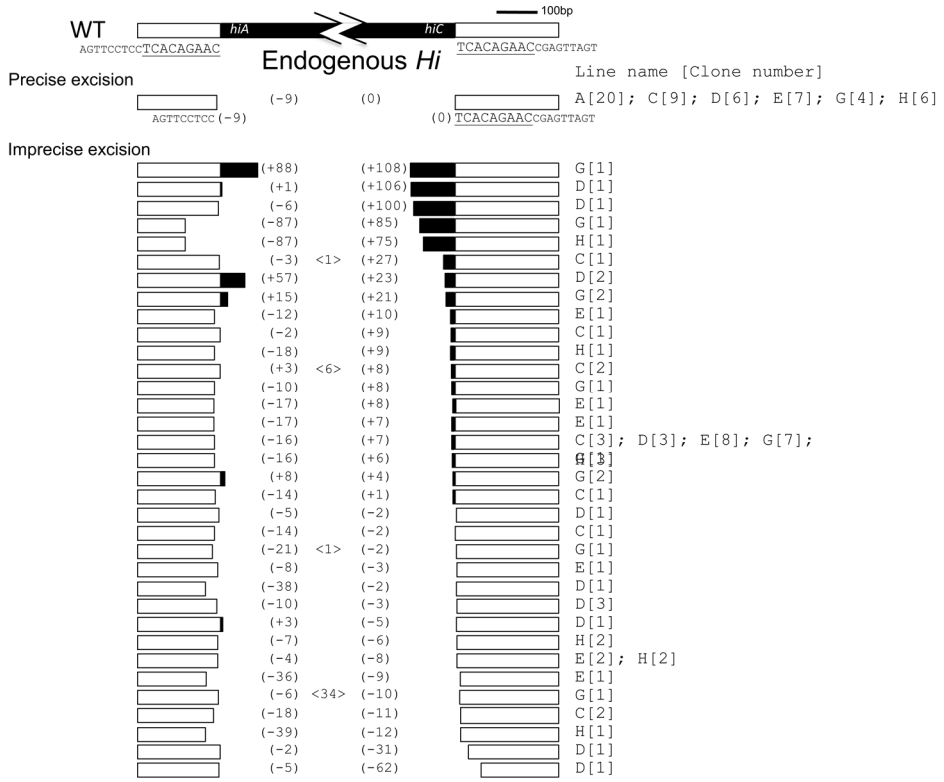


Figure 7

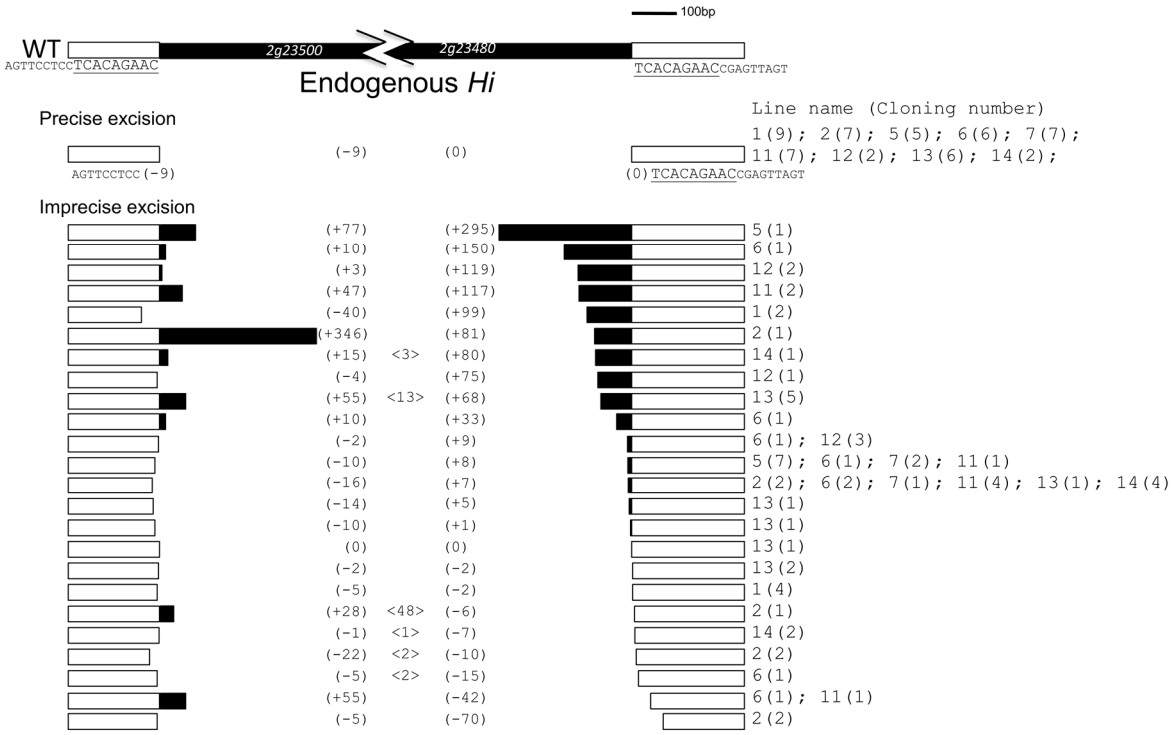
Excision of *Hi* in *ddm1* plants detected by PCR. Genomic DNA of 11 *ddm1* plants (lane numbers from 2 to 12) and 5 wild type sibling plants (lane numbers from 13 to 17) were used to analyze excision of endogenous *Hi* by nested PCR. These lines are derived from segregating population in self-pollinated progeny of a *DDM1/ddm1-1* heterozygote. Positions of primers used in PCR are shown in top and their sequences are shown in primer Table.

Excision pattern of *Hi* in *ddm1* and *Hi* TG lines

A



B



(Continued on the next page.)

Figure 8

Excision product patterns at endogenous *Hi* locus in *Hi* TG and *ddm1* lines.

(A), (B) Excision patterns of endogenous *Hi* at *ddm1* lines (top panel) and *Hi* TG lines (bottom panel). 120 clones of excision PCR products from 6 *ddm1* plants and 116 clones from 9 transgene plants were sequenced. Excision patterns were grouped to two groups. Precise excision, with the intact flanking sequence but except one TSD sequence. By contraries, imprecise excision, all excision patterns but are not precise excision pattern. White bar indicates flanking sequence of endogenous Hi. Black bar indicates endogenous Hi sequence. TSDs are emphasized by underlines. Number with plus sign in () indicate how long terminal sequences of endogenous *Hi* were detected. Number with minus sign in () indicate how long flanking sequence were deleted.

Schematic diagram for transgenes of *Hi*

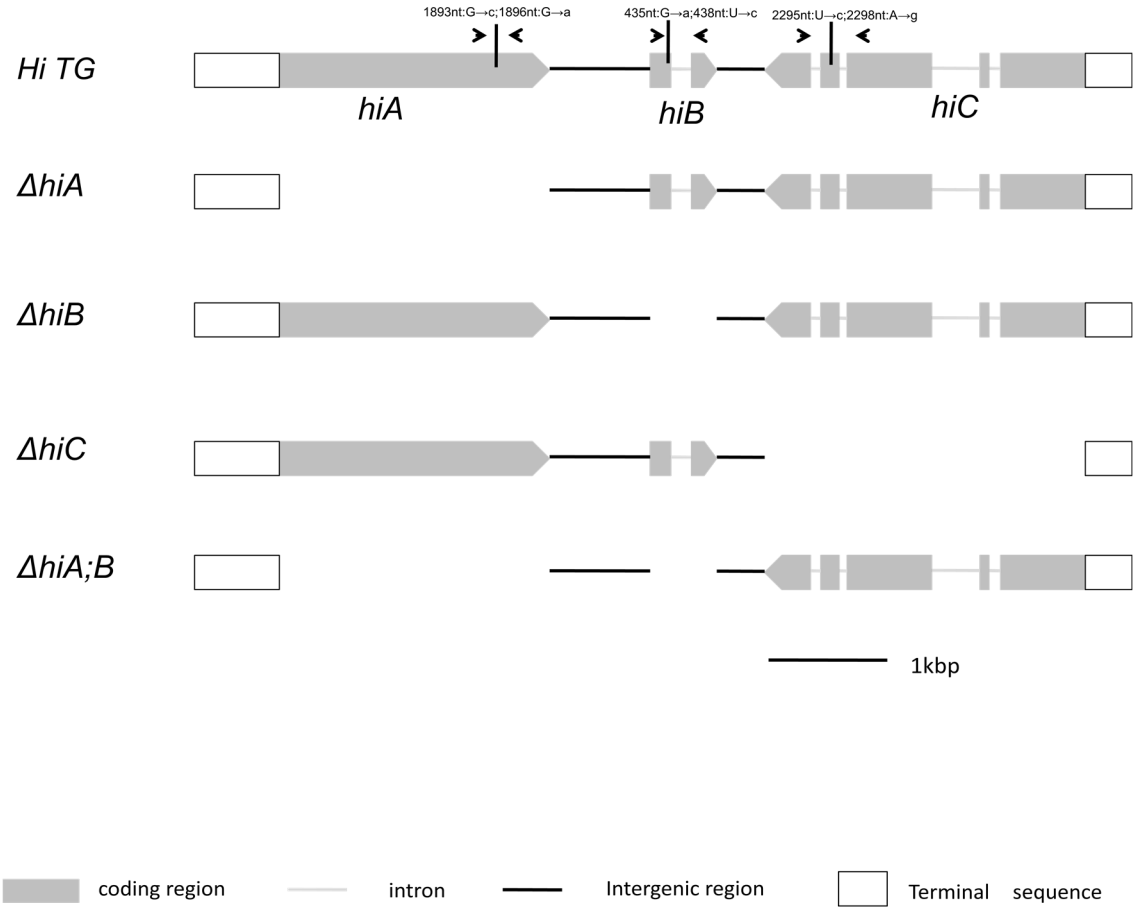


Figure 9

Schematic diagram for transgenes with deletion in each of the three ORFs. The deletion constructs, *Δ hiA*, *Δ hiB*, *Δ hiC* and *Δ hiA;B* have two silent mutations in each of the remaining ORFs. . Primer sets used at Reverse Transcription Polymerase Chain Reaction (RT-PCR) for detecting transcriptional products from *Vandal21* subfamily are exhibited by arrow heads.

Excision of endogenous *Hi* in *Hi* TG lines

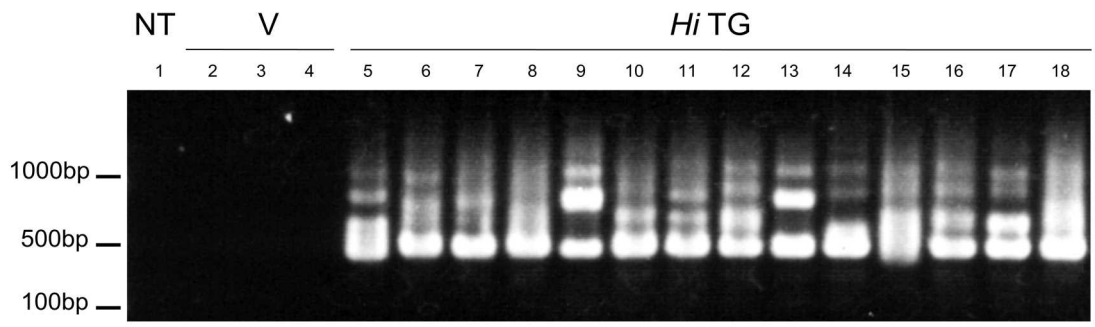


Figure 10

Introduction of *Hi* transgene induces excision of endogenous *Hi* copy. Excision of endogenous *Hi* induced by transgene for *Hi* (*Hi* TG: lanes 5-18). Lanes 1 and 2-4 are control nontransgenic (NT) and transfomant lines with empty vector (V) used for negative controls, respectively.

Excision of endogenous *Hi* in the sibling lines of *Hi* TG

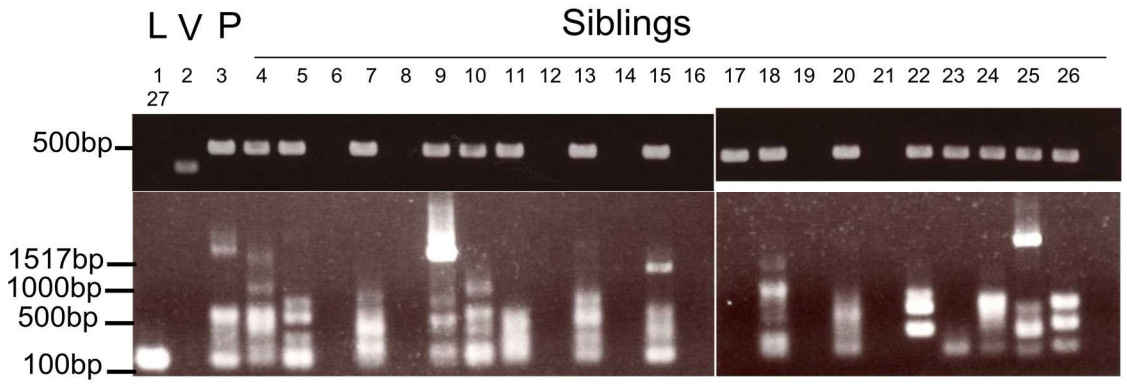


Figure 11

Top panel; Genotyping of *Hi* transgene. Genomic DNA of first generation of *Hi* transgene line (P; lane3) and the siblings of the transgene line were used (siblings; lane 4-27). A Wild Type line (Ecotype Landsberg; L; lane1) and an empty vector line (V; lane 2) were used as negative control.

Bottom panel; Excision activity of endogenous copy is keeping by *Hi* transgene. Genomic DNA of WT (L; lane1) and empty vector (V; lane 2) were used as control. Genomic DNA of first generation of *Hi* transgene line (P; lane3) and the siblings of the transgene line were used (siblings; lane 4-27). In this experiment, used primer sets were unique to other excision PCR (see primer table).

Changes of DNA methylation at endogenous *Hi* terminal regions

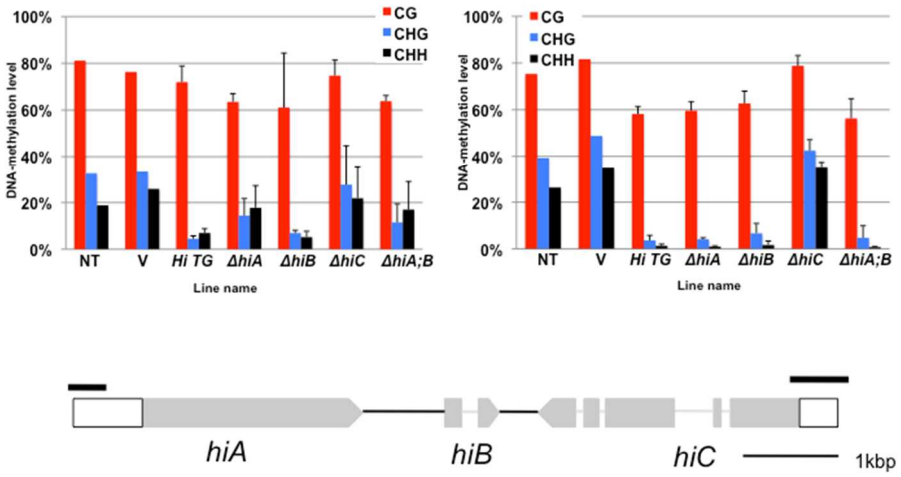


Figure 12

Bisulfite sequencing method was used at analysis of DNA-methylation level change at both terminal sequence of endogenous *Hi*. One line of WT, one transgenic line of empty vector, four transgenic lines of *Hi* TG, four transgenic lines of Δ*hiA*, four transgenic lines of Δ*hiB*, four transgenic lines of Δ*hiC* and four transgenic lines of Δ*hiA;B* were used at bisulfite sequencing. The black bars exhibit the regions for bisulfite sequencing. Methylation status of *Hi* termini in the transgenic line and progeny

Changes of DNA methylation in the sibling lines of *Hi* TG

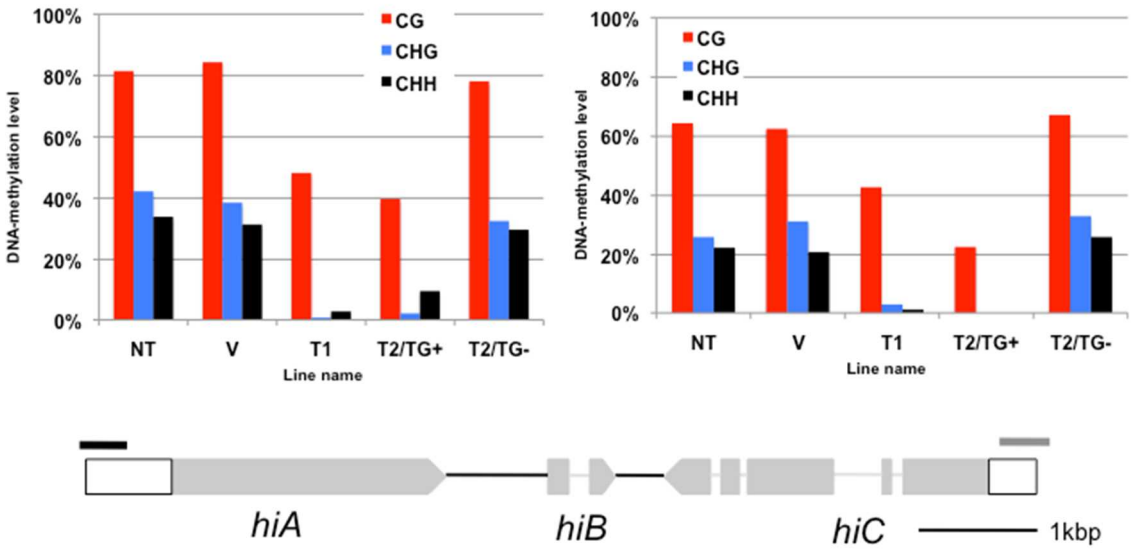


Figure 13

Methylation status of *Hi* termini in the transgenic line and progeny. T1 transformant with *Hi* transgene showed reduction of DNA methylation in both termini, compared to non-transgenic (NT) and transformants with empty vector (V). T2/TG- and T2/TG+ are self-pollinated progeny of the T1 without and with transgene, respectively. Regions upstream of *hiA* and *hiC* were examined as shown in the bottom.

Excision of endogenous *Hi* in several transgenic lines

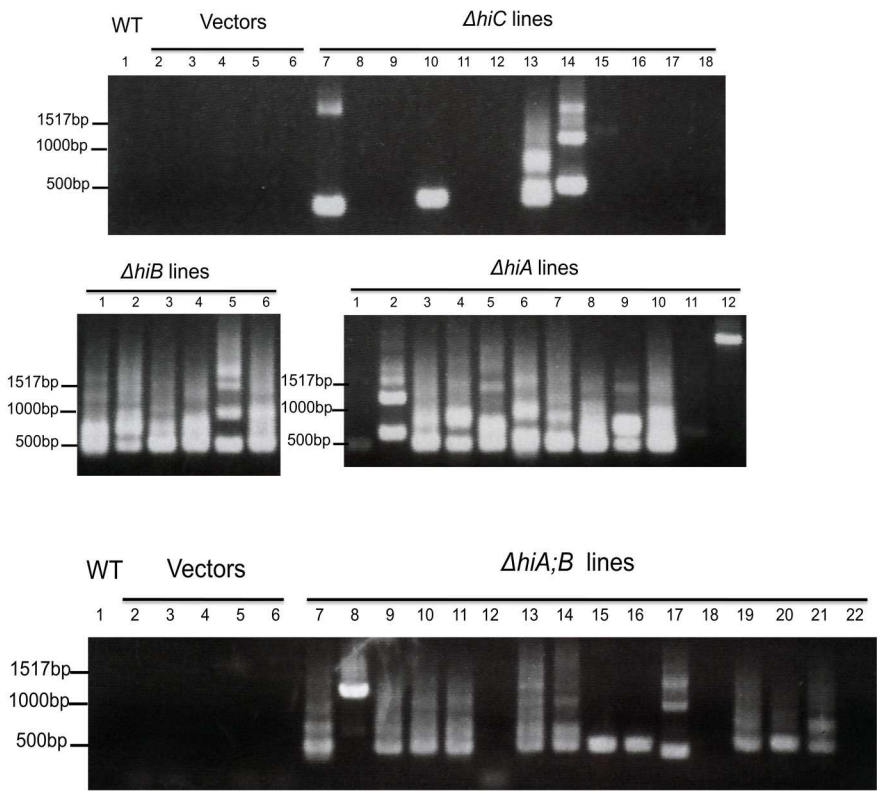


Figure 14

Excision of endogenous *Hi* in Δ*hiC* lines (top), which were induced by transgene (Δ*hiC* lines: lanes 7-18). Lanes 1 and 2-6 are Wild Type (WT) and transformant lines with empty vector (V) used for negative controls, respectively. The negative controls are common to Δ*hiB* and Δ*hiA* lines. Excision of endogenous *Hi* in Δ*hiB* lines (left of middle), and in Δ*hiA* lines (right of middle).

Excision of Δ*hiA;B* lines (bottom). Negative controls (WT, lane 1; V, lane2-6), and Δ*hiA;B* lines (lane 7-22).

Transcriptional activity of endogenous *Hi* in Δ*hiA* lines

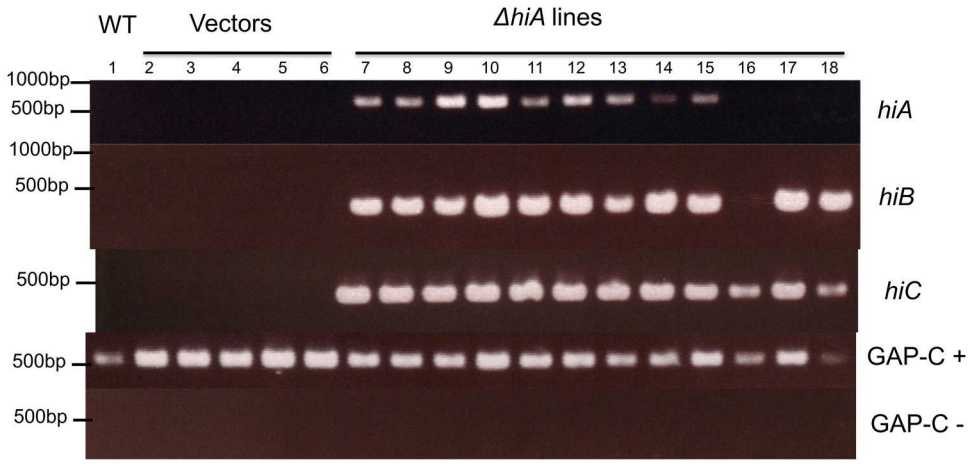


Figure 15

RT-PCR method was used for detecting transcriptional products of coding genes of *Hi* in Δ*hiA* transgenic lines (lane numbers from 7 to 18). WT (lane numbers 1) and empty vector transgenic lines (lane numbers from 2 to 6) were used for negative control.

Transcriptional activity of endogenous *Hi* in several transgenic lines

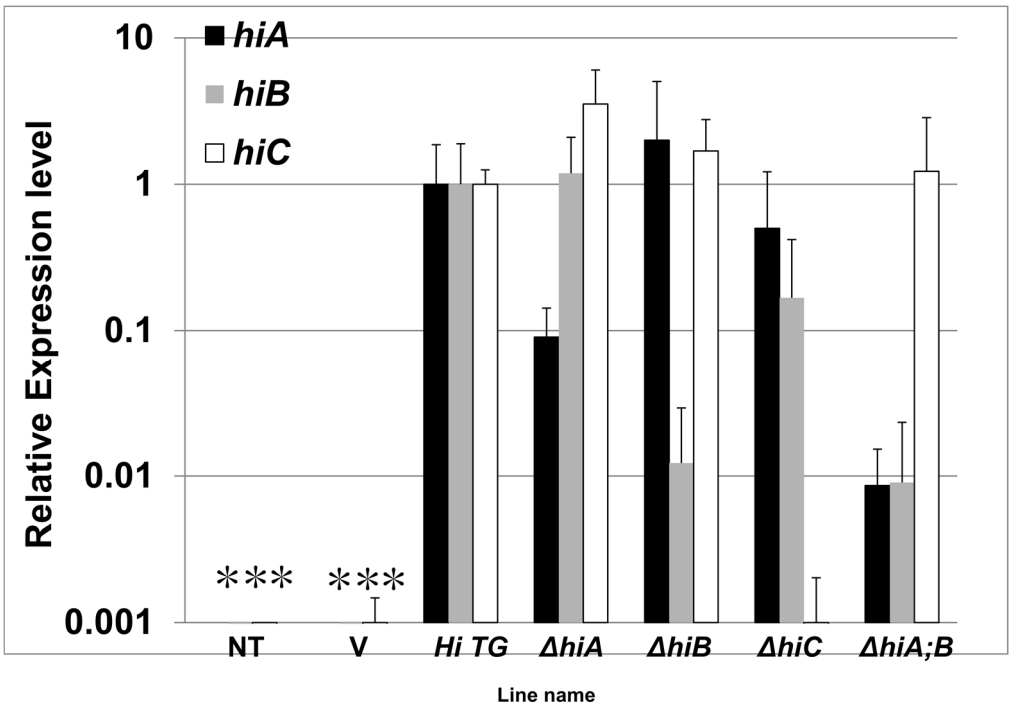


Figure 16

Transcript levels of three ORFs measured by quantitative RT-PCR. For each of the ORFs, the average value was normalized by the average value of full length *Hi*. I examined three transgenic lines for empty vector, four lines for full length *Hi*, 12 lines for Δ*hiA*, six lines for Δ*hiB*, 12 lines for Δ*hiC*, and 18 lines for Δ*hiAB*. Each bar indicates standard deviation among the values for different transgenic lines. Asterisks indicate average less than 0.001. Independent detections of the signal using the same RNA samples are shown in Figure 17, 18, 19.

Transcriptional activity of endogenous *Hi* in Δ*hiB* lines

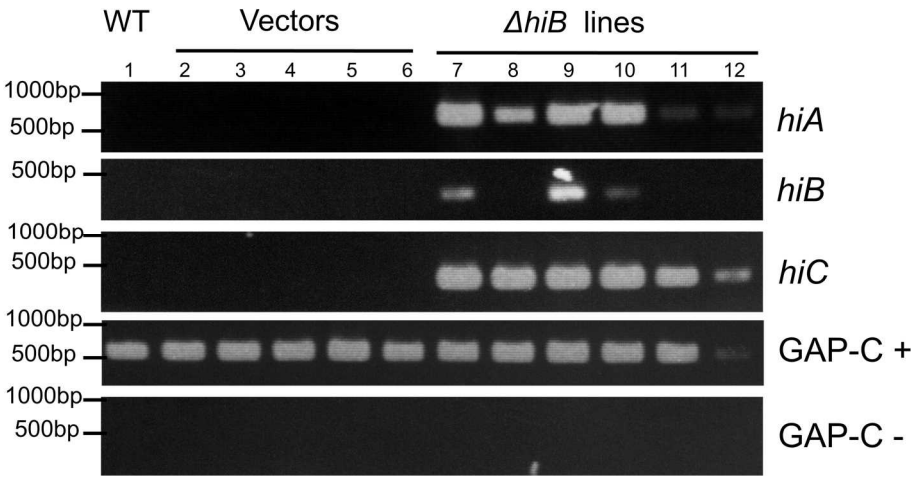


Figure 17

RT-PCR method was used for detecting transcriptional products of the coding genes of *Hi* in Δ *hiB* lines (lane numbers from 7 to 12). WT (lane numbers 1) and empty vector transgenic lines (lane numbers from 2 to 6) were used for negative control.

Transcriptional activity of endogenous *Hi* in Δ*hiC* lines

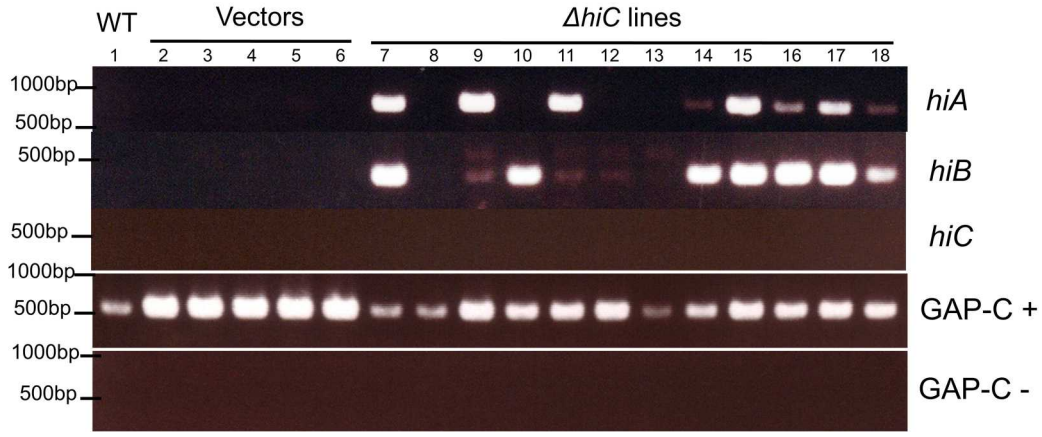


Figure 18

RT-PCR method was used for detecting transcriptional products of the coding genes of *Hi* in Δ *hiC* lines (lane numbers from 7 to 18). WT (lane numbers 1) and empty vector transgenic lines (lane numbers from 2 to 6) were used for negative control.

Transcriptional activity of endogenous *Hi* in Δ*hiA;B* lines

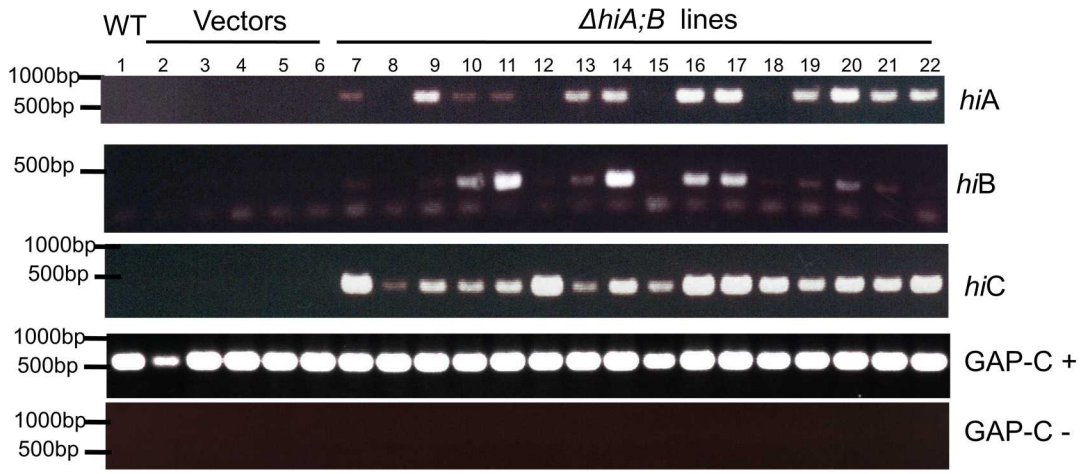


Figure 19

RT-PCR method was used for detecting transcriptional products of the coding genes of *Hi* in Δ *hiA;B* lines (lane numbers from 7 to 22). WT (lane numbers 1) and empty vector transgenic lines (lane numbers from 2 to 6) were used for negative control.

Transcriptional activity of endogenous *Hi* in the sibling lines of *Hi* TG

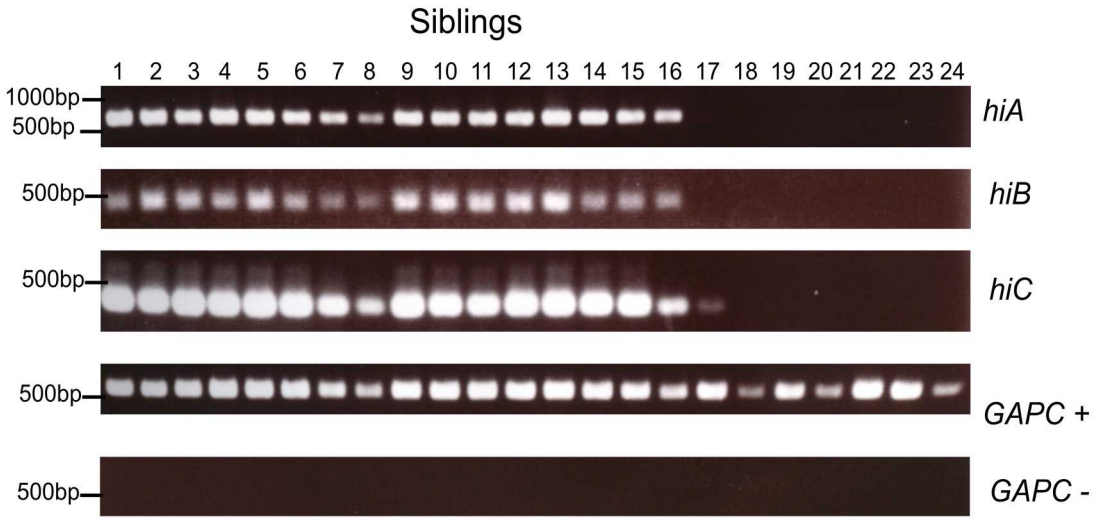


Figure 20

RT-PCR method was used for detecting transcriptional products of coding genes of *Hi* at the siblings with transgene (lane 1-16) and without transgene (lane 17-24).

Southern plot of T_1 and T_2 generation of *Hi* TG lines

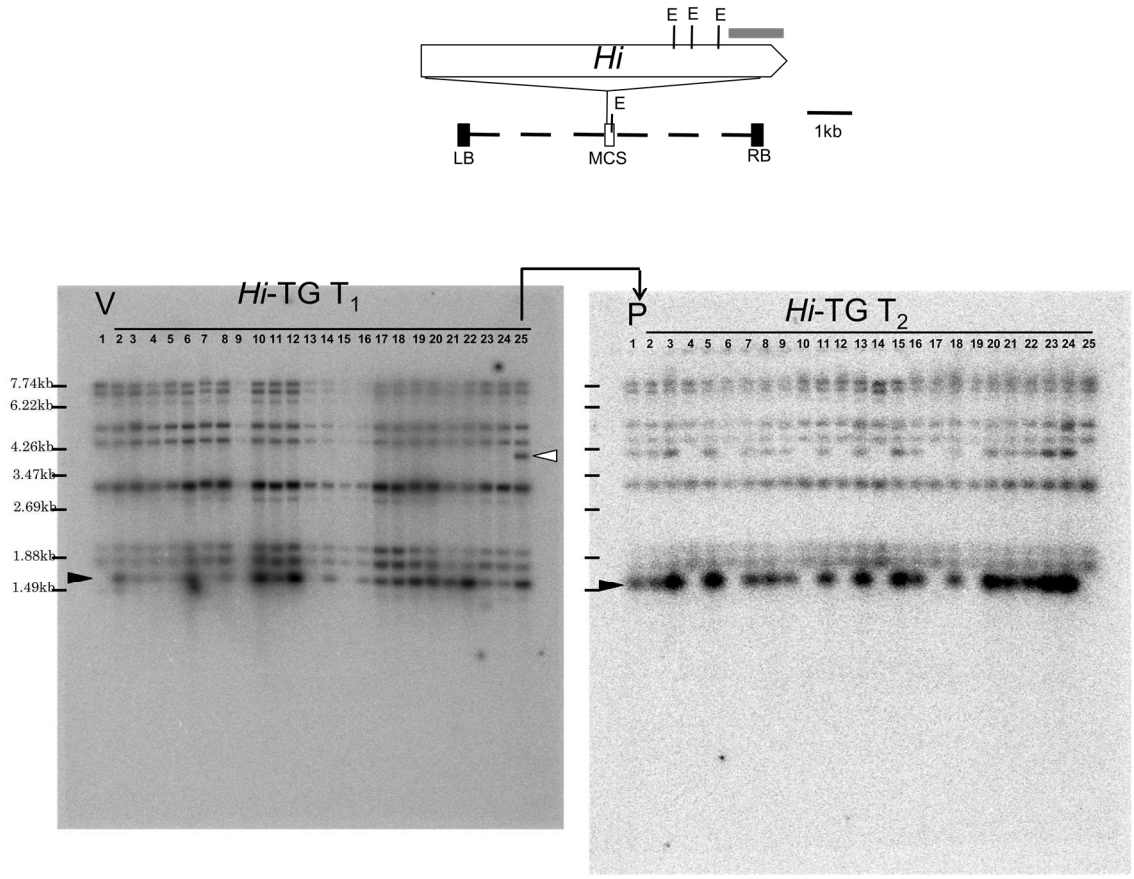


Figure 21

Genomic DNA of transgenic plants were digested with EcoRV. White triangle indicates band of putative duplicated copy of *Vandal21*. Black trangle indicates bands originated from *Hi* transgene. Arrow indicates parent plant of T_2 generation of full-length *Hi* transgenic plants. Grey bar shows probe used at southern bolt. Note the probe not specific to *Hi*, but some copies of *Vandal21*. vec: empty vector transgenic plan (V; lane1 left) ; MCS LB and RB: multi-cloning sites, left and right border of vector; E: EcoRV site; parent plants (P; lane:1 right).

Excision of endogenous *Hi* and *Hi* TG transgene in *Hi* TG lines

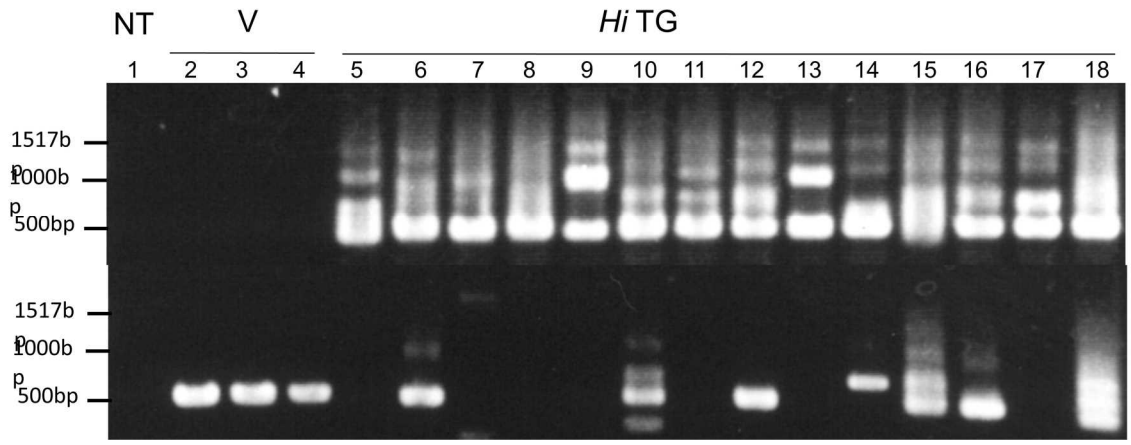


Figure 22

Excision of endogenous *Hi* (top panel) and *Hi* transgene (bottom panel) which were induced by *Hi* transgene (*Hi* TG: lanes 5-18). Lanes 1 and 2-4 are non-transgenic (NT) and transformant lines with empty vector (V) used for negative controls, respectively.

Materials and Methods

Plant Materials Col ecotype was used as Wild Type (WT) plant. The *ddm1-1* mutant allele was used throughout. Details for self-pollination of *ddm1*/*ddm1* mutants and wild type *DDM1*/*DDM1* siblings were described previously (Kakutani et al 1996).

Identification of mobile *VANDAL21* copies To detect *de novo* insertions of *Vandal21*, suppression PCR and whole genome re-sequencing were used.

In suppression PCR, genomic DNA was digested by blunt end restrict enzyme, firstly. And then the samples were ligated with an adaptor. Finally, nested PCR was performed with the following conditions. First PCR: 95°C for 2min, 25 cycles of (95°C for 30 sec; 68°C for 5min), and 72°C for 2min. The products were diluted to 1000 times by H2O, and were used for second PCR: 95°C for 2min, 25 cycles of (95°C for 30 sec; 68°C for 5min), and 72°C for 2min.

Whole genome re-sequencing libraries were constructed by Paired End DNA Sample Prep Kit (Illumina). The genomic libraries were sequenced on the Illumina Genome Analyzer IIx sequencers. Sequences and quality scores passing through the standard Illumina pipeline filters were retained for further analysis. Reads of Whole-genome sequencing were filtered by 35 nucleotides (nts) of outmost terminal sequences of 7 *VANDAL21* copies. The downstream 20 nts of *Vandal21* terminal sequences were used to database alignment. Reads are classed to *de novo* insertions of *Vandanl21*, if it was perfect matching to unique locus of *A. thaliana* genome.

In both approaches, each of the *VANDAL21* copies can be distinguished by polymorphisms in the terminal regions.

Characterization of excision of endogenous *Hi* To detect excision of *Hi*, I used nested PCR with the following conditions. First PCR: 94°C for 2min, 25 cycles of (94°C for 30 sec; 55°C for 30 sec; 72°C for 30 sec), and 72°C for 2min. The products diluted to 20 times by H2O was used for second PCR: 94°C for 2min, 27 cycles of (94°C for 30 sec; 55°C for 30 sec; 72°C for 30 sec), and 72°C for 2min. The PCR products of excision were identified by TA-cloning and

sequencing.

Construction of transgenes Full length *Hi* was recovered by PCR and cloned into vector PZP2H-lac (Fuse et al. 2001). In most constructs, I introduced silent mutations for each of the three ORF, so that transcripts from the transgene and endogenous copies could be distinguished between. Transgenes with deletion in each of the three ORFs were generated by PCR using the full length *Hi* as the template.

Transcriptional analysis Total RNA was isolated by Promega SV Total RNA Isolation System (cat. # Z3100). I did reverse transcription reaction by using Takara RNA PCR Kit (RR019A), following manufacturer's instructions. Oligo dT-Adaptor primer were chosen to reverse transcribe transcriptional products of 3 coding genes of *Hi*. GAPC was used for control. I must note that these primer sets are not only amplifying *Hi*, but also amplifying some copies of *Vandal21*. For distinguishing mRNA originating from endogenous or transgenic *Hi*, direct sequencing was used.

qRT-PCR were performed by using SYBR Premix Ex Taq □(TaKaRa) on Takara Thermal Cycler Dice_Real Time System TP800(TaKaRa). The cycling conditions comprised 95°C for 30 sec and 40 cycles at 95°C for 5 sec, 60°C for 30 sec and 72°C for 30 sec. The gene *UBC* (*At5g25760*) was used as in internal control. One WT plant, three plants of empty vector, four plants of *Hi* TG lines, twelve plants of Δ*hiC* lines, six plants of Δ*hiB* lines, twelve plants of Δ*hiA* lines and sixteen plants of Δ*hiA;B*, each with 3 technical replicates, were averaged. We must note primers are not specific to *Hi*, but some *Vandal21* subfamily copies.

Reference

Benito MI., Walbot V. (1997). Characterization of the maize Mutator transposable element MURA transposase as a DNA-binding protein. *Mol. Cell Biol.* **17**: 5165-5175.

Brown J., Sundaresan V. (1992). Genetic study of the loss and restoration of Mutator transposon activity in maize: evidence against dominant-negative regulator associated with loss of activity. *Genetics* **130**: 889-898.

Brown WE., Robertson DS., Bennetzen JL. (1989). Molecular analysis of multiple mutator-derived alleles of the bronze locus of maize. *Genetics* 122: 439-445.

Chalvet F., Grimaldi C., Kaper F., Langin T., Daboussi MJ. (2003). Hop, an active Mutator-like element in the genome of the fungus Fusarium oxysporum. *Mol. Biol. Evol.* **20**: 1362-1375.

Chandler V.L., Walbot V. (1986). DNA modification of a maize transposable element correlates with loss of activity. *Proc. Natl. Acad. Sci. USA* **83**: 1767-1771.

Cui H., Fedoroff NV. (2002). Inducible DNA demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein. *Plant Cell* **14**: 2883-2899.

Dietrich CR., Cui F., Packila ML., Li J., Ashlock DA., Nikolau BJ., Schnable PS. (2002). Maize Mu transposons are targeted to the 5' untranslated region of the gl8 gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics* **160**: 697-716.

Eisen JA, Benito MI, Walbot V. (1994). Sequence similarity of putative transposases links the maize Mutator autonomous element and a group of bacterial insertion sequences. *Nucleic Acids Res.* **22**: 2634-2636.

Fuse T., Sasaki T., Yano M. (2001). Ti-plasmid vectors useful for functional analysis of rice genes. *Plant Biotechnology* **18**: 219-222.

Gao D. (2012). Identification of an active Mutator-like element (MULE) in rice (Oryza sativa). *Mol. Genet. Genomics* **287**: 261-271.

Hardeman KJ., Chandler VL. (1989). Characterization of bz1 mutants isolated from mutator stocks with high and low numbers of Mu1 elements. *Dev. Genet.* **10**: 460-472.

Hoen DR., Park KC., Elrouby N., Yu Z., Mohabir N., Cowan RK., Bureau TE. (2006). Transposon-mediated expansion and diversification of a family of ULP-like genes. *Mol. Biol. Evol.* **23**: 1254-1268.

Kakutani T., Jeddeloh JA., Flowers SK., Munakata K., Richards EJ. (1996). Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl. Acad. Sci. USA* **93**: 12406-12411.

Kato M., Miura A., Bender J., Jacobsen SE., Kakutani T. (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Curr. Biol.* **13**: 421-426.

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

Lippman Z., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 810-813.

Liu S., Yeh CT., Ji T., Ying K., Wu H., Tang HM., Fu Y., Nettleton D., Schnable PS. (2009). Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genetics* **5**: e1000733.

Lisch D. (2002). Mutator transposons. *Trends Plant Sci.* **7**: 498-504.
Lisch D., Girard L., Donlin M., Freeling M. (1999). Functional analysis of deletion derivatives of the maize transposon MuDR delineates roles for the MURA and MURB proteins. *Genetics* **151**: 331-341.

Marquez CP., Pritham EJ. (2010). Phantom, a new subclass of Mutator DNA transposons found in insect viruses and widely distributed in animals. *Genetics* **185**: 1507-1517.

Martienssen R. (1996). Epigenetic phenomena: paramutation and gene silencing in plants. *Curr. Biol.* **6**: 810-813.

Martienssen R., Baron A. (1994). Coordinate suppression of mutations caused by Robertson's mutator transposons in maize. *Genetics* **136**: 1157-1170.

McClintock B. (1951). Chromosome organization and genic expression. *Cold Spring Harbor Symp. Quant. Biol.* **16**: 13-47.

McClintock B. (1958). The suppressor-mutator system of control of gene action in maize. *Carnegie Institution of Washington Year Book* **57**: 415-429

Mirouze M., Reinders J., Bucher E., Nishimura T., Schneeberger K., Ossowski S., Cao J., Weigel D., Paszkowski J., Mathieu O. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427-430.

Miura A., Yonebayashi S., Watanabe K., Toyama T., Shimada H., Kakutani T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature* **411**: 212-214.

Schläppi M., Raina R., Fedoroff N. (1994). Epigenetic regulation of the maize Spm transposable element: novel activation of a methylated promoter by TnpA. *Cell* **77**: 427-437.

Schläppi M., Raina R., Fedoroff N. (1996). A highly sensitive plant hybrid protein assay system based on the Spm promoter and TnpA protein for detection and analysis of transcription activation domains. *Plant Mol. Biol.* **32**: 717-725.

Singer T., Yordan C., Martienssen RA. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes Dev.* **15**: 591-602.

Tsukahara S., Kobayashi A., Kawabe A., Mathieu O., Miura A., Kakutani T. (2009). Bursts of retrotransposition reproduced in Arabidopsis. *Nature* **461**: 423-426.

Xu Z., Yan X., Maurais S., Fu H., O'Brien DG., Mottinger J., Dooner HK. (2004). Jittery, a Mutator distant relative with a paradoxical mobile behavior: excision without reinsertion. *Plant Cell* **16**: 1105-1114.

Yu Z., Wright SI., Bureau TE. (2000). Mutator-like elements in Arabidopsis thaliana. Structure, diversity and evolution. *Genetics* **156**: 2019-2031.

Acknowledgments

First of all, I would like to give my gratitude to Dr. Tetsuji Kakutani. He gave me opportunity and appropriate direction to this study. And I also would like to thank all members of Kakutani's laboratory. They gave me much helpful advises to this work. Finally, I would like to have a special thanks to my parents. If it had not been for their support, I would never have finished this work.

Target	Experiment	Figure	Forward primer sequence	Reverse primer sequence
Endogenosu *Hi*	*Hi* TG cloning	Fig9	TACGGGCCCGAATAATCGTCTGGCCAGTCCCTT	ATCGTCGACGAGGGATCATCTCTTGTGTCCCT
At2g23500(*hiA*)	introducing silent mutation to *Hi* TG	Fig9	AGTGGTCGAACTAAACTCATTCGAGCGTGA	TTTAGTTCGACCACTTTCAGCTTCTCGGCA
At2g23490(*hiB*)	introducing silent mutation to *Hi* TG	Fig9	AAGGGAAAGCGTTGATTACAATCGGAAGA	TCAACGCTTTCCCTTAAGCTACTCACCTCT
At2g23480(*hiC*)	introducing silent mutation to *Hi* TG	Fig9	CAACCACGTGCTCGGAACAACTGAGGTTAG	TCCGAGCACGTGGTTGATTTGCTCAAGGGT
Hi	constructing *ΔhiA* transgene	Fig9	GCAGATTACAGTTTTTAACTTTGTTTCTGC	AGTTAAAAACTGTAATCTGCCAAAACAATA
Hi	constructing *ΔhiB* transgene	Fig9	TCTCTCACATTGTGTTATCCTATTGTTCCT	GGATAACACAATGTGAGAGAATTCGAGTCG
Hi	constructing *ΔhiC* transgene	Fig9	ATATTACCAAGACTGATTTCGAATCGGAAA	GAAATCAGTCTTGGTAATATCGCGTAATAC
Hi	constructing *ΔhiA;B;C* transgene	Fig9	GCAGATTACAGACTGATTTCGAATCGGAAA	GAAATCAGTCTGTAATCTGCCAAAACAATA
Hi	constructing *ΔhiA;B* transgene	Fig9	TCTCTCACATTGTGTTATCCTATTGTTCCT	GGATAACACAATGTGAGAGAATTCGAGTCG
Hi	constructing *ΔhiA;C* transgene	Fig9	ATATTACCAAGACTGATTTCGAATCGGAAA	GAAATCAGTCTTGGTAATATCGCGTAATAC
Hi and adaptor of supression PCR	suppression PCR (first PCR)	Fig1	GGATCCTAATACGACTCACTATAGGGC	CAAAGCTTTTGAAGCTCTCTCCATACC
Hi and adaptor of supression PCR	suppression PCR (second PCR)	Fig1	AATAGGGCTCGAGCGGC	GCTTGCAGGAGGAGAAAAACGACAATG
At2g13290	Identifying Target Site Duplication	Fig6	GATTAAGAAATGAGAACACACG	CTGGAAAACATCATGACCTTA
At2g23450	Identifying Target Site Duplication	Fig6	TGCGAAATAACAATCAGAGTA	GATATCCCAATTGCTCGTTGA
At2g23830	Identifying Target Site Duplication	Fig6	GTTCCATGTTGAATAATCAGC	GACGCTTATCCGCATAGTTCT
At3g30851	Identifying Target Site Duplication	Fig6	GTTTTGAAATCGAAGAGAGC	GGACATTTTAGCGACTAAACT
At3g32111	Identifying Target Site Duplication	Fig6	AGAAAGCTGGAGAGGCTAATG	TCCATCAACCACCGTTCTGGT
At5g33405	Identifying Target Site Duplication	Fig6	TCAATTAGGCAATTGAGCACT	AAGTGAAGAGATAGATCGATT
At4g04720	Identifying Target Site Duplication	Fig6	GTAAAGGAGGAGACTTTCGTT	TCAGCAAATTGACAAAGACA
flanking sequences of *Hi*	excision PCR(first PCR)	Fig7; 10; 19; 21	ACGAGCAGAAAACATGCCACCA	TGCTCTAAACATTGCCTGAAGC
flanking sequences of *Hi*	excision PCR(second PCR)	Fig7; 10; 19; 21	CGACGAGCTACGTTACTGGG	AGTCTATTCACCATCGCCTAGTT
flanking sequences of *Hi*	excision PCR(first PCR)	Fig11	CGACGAGCTACGTTACTGGG	AGTCTATTCACCATCGCCTAGTT
flanking sequences of *Hi*	excision PCR(second PCR)	Fig11	AGTCGAGAGCTTTGATTCGTTGA	GCCAAGTGTGTAAGGCCCAT
At2g23500(*hiA*)	RT-PCR	Fig14;16;17;18	CAGGAGTTAAGTCGGGTCTAC	TGCGACCTATCCGGAACAAGA
At2g23490(*hiB*)	RT-PCR	Fig14;16;17;18	GACCCCTACTACGATGATATG	CCATAGGATTACGGAATACCA
At2g23480(*hiC*)	RT-PCR	Fig14;16;17;18	ACAGCTGTGGGAACTTCCTCT	AACACTCAGTCACCATGGCCT
At3G04120(*GAPC*)	RT-PCR	Fig14;16;17;18	CACTTGAAGGGTGGTGCCAAG	CCTGTTGTCGCCAACGAAGTC
At2g23500(*hiA*)	qRT-PCR	Fig15	GATGGTGCCTTTGGTCGAGA	TTTCAAAAGCAAGCTCACCGT
At2g23490(*hiB*)	qRT-PCR	Fig15	TAGCATTGTCGAGACGCGAA	ATCCCAAAGTTTACGGATGTGC
At2g23480(*hiC*)	qRT-PCR	Fig15	AGGATGTGCAAGGTGAGTTTCA	ACTCCCGTGATTTCAGCCAA
At5g25760(*UBC*)	qRT-PCR	Fig15	CTGCGACTCAGGGAATCTTCTAA	TTGTGCCATTGAATTGAACCC

Terminal sequence of *At2g23480*	Bisulfite-sequencing	Fig12	CTTTCTTCRCCRRCACCTTCTCCTTCACTTTCTCA	ATGGGTATTGAAAAAGTYGAGAGYTTTGATTYGTTG
Terminal sequence of *At2g23480*	Bisulfite-sequencing	Fig13	GGAAATAGAAYGGAAATAYTGGTGAAAGYAAGAAG	CTACTTCTTCTTCCTTCTCTRTAACTACTAACTC
Terminal sequence of *At2g23500*	Bisulfite-sequencing	Fig12; 13	AATCTCAACATCCTCAAAATATRTAATTCAAARCT	GTTAGAAGAAAAAAAAAYTAAATGGGYYAAGTGTGT

Primer table

Primer table

Hi 3'	ACAAAAAAGGGACTGGCCAGACGATTATTC
Hi 5'	CCACAAAAGGGACACAAGAGATGATCCCTC
AT4TE15615 5'	GAATAATAATCTCTTGTGTCCCTTTTTTGG
AT4TE15615 3'	CCAAAAAAGGGACACAAGAGATTATTATTC

Whole genome Resequencing

Integration number	Line_name	Read_number	Terminal_sequence	Flanking_sequence	Integration site
1-1	A2	7	*Hi* 3'	TTTTCCTCAAGGGCTTATTC	*At2g23450* upstream
1-2	A2	38	*Hi* 3'	AAGTCAGTGTAAGATTGATT	*At2g23370* intron
2-1	C2	2	*Hi* 3'	TTTGCAAGATCTACTAAATT	*At1g11655* upstream
2-2	C2	1	*Hi* 3'	TATTTAAAGCTTGTGAATTC	*At2g11990* uptream
2-3	C2	13	*Hi* 3'	GTTTGAAGGCCTAGGTCACT	*At4g07920*
2-4	C2	2	*Hi* 3'	ATTTAGATGCTAGGTGATTT	*At2g04305* downstream
2-5	C2	8	*Hi* 3'	CAAGTACGCGGTCGAGTGAAACGGTCGAGTATA	*At4g06517* upstream
3-1	H4	2	*Hi* 3'	TTCGGATGGTTTGGTTCGGA	*At2g23470* upstream
3-2	H4	10	*Hi* 3'	GTTTTCTTTCGATAAAGATC	*At2g13280* upstream
3-3	H4	1	*Hi* 3'	GATGAGGGAGAAGAGAGATG	*At2g23490* upstream
3-4	H4	24	*Hi* 3'	CTCGGAAGGAGAAGGGAAAT	*At2g23830* upstream
3-5	H4	15	*Hi* 3'	CATACGATTAAAGAAACCCA	*At2g42780* upstream
3-6	H4	2	*Hi* 3'	ACTTCAGTCACACGTTTTTC	*At2g32400* upstream
3-7	H4	4	*Hi* 3'	AATTAAAAGAATCCCTGGAC	*At3g10670* upstream
3-8	H4	1	*Hi* 3'	TTTCGAGTCGAATATGACTTGTTGTCAT	*At5g31804-At5g31821*
4-1	J1	2	*Hi* 3'	TTCATTCCGTTAGAAAAGTG	*At1g37110* upstream
4-2	J1	1	*Hi* 3'	TTCAAAAAACCCACACACAC	*At5g51480* upstream
4-3	J1	3	*Hi* 3'	TAAATAACATCAACATGCAT	*At2g24240* upstream
4-4	J1	1	*Hi* 3'	GTTCCTTTCCCTCTCTGAAA	*At4g31530* upstream
4-5	J1	1	*Hi* 3'	GAACCGAGTTAGTAGTTACA	*At2g23470* upstream
4-6	J1	5	*Hi* 3'	CTTTTGGAGGTTTGTGGAGA	*At5g34930* upstream
4-7	J1	1	*Hi* 3'	CGGGATCCACTTTGTTATAAACCTAAGTATCTGCAATTAGGGTTGTTGCTACTTCAC	*At1g37057*
4-8	J1	2	*Hi* 3'	AAACATAACATGAGTGTACG	*At1g12070*
4-9	J1	2	*Hi* 5'	GTGCGACTGTTACAGTAACA	*At5g11880* upstream
4-10	J1	2	*Hi* 5'	GAAGGAGGGGAAGGCAGGGA	*At5g12150* upstream
5-1	L2	1	*Hi* 3'	GTTTTCGTAGCCATGGCTTC	*At2g05755* upstream

5-2	L2	2	*Hi* 3'	GTCGGTCGACACTGTCGCGTTCTGTAGTGTCTACCGTGGCTGG	*At5g32386*
5-3	L2	2	*Hi* 3'	GCCTTAGCAAGTGCATCGGC	*At2g11240*
5-4	L2	35	*Hi* 3'	CCTTTAGTCGGACCGAACAC	*At3g32110*
5-5	L2	1	*Hi* 3'	AGTGGAAACGCTACAGAGTT	*At5g37010*
5-6	L2	22	*Hi* 3'	ACTTCAGTCACACGTTTTTC	*At2g32400* upstream
5-7	L2	10	*Hi* 3'	AATAAAAAGAGTAAGAGAAG	*At5g53170* upstream
5-8	L2	1	*Hi* 5'	TTCCACTCTCTCAAGGTGAT	*At5g37010*
5-9	L2	1	*Hi* 5'	TCTGTGAGAATAATCGTCTG	*At2g23470* upstream
5-10	L2	1	*Hi* 5'	CTAAAGGTGACTATGAAAAG	*At3g32110*
AT4TE15615-1	A2	1	AT4TE15615 5'	TATAGCTCAAATCGCCTTAA	*At2g09890* downstream
AT4TE15615-2	C2	1	AT4TE15615 5'	TGATCTTTCCCCTTCTCTGACAAGGATCA	*At4g06550*
AT4TE15615-3	J1	1	AT4TE15615 5'	GATATGAAGACTACAAAAGGATTTTACA	*At4g06552* upstream

Suppression PCR

Integration number	Line_name	Terminal_sequence	Flanking_sequence	Integration site
1-4	A2	*Hi* 5'	TGAGGAAAAGCGTGGGAAGATTAAAATGACGGATATGTCCTTATCATGGT	*At2g23450* upstream
1-5	A2	*Hi* 5'	CTCCTTTCACCACCAATCTCTCGTTTCATTTACTACAACTACAACACTCT	*At2g23520* upstream
2-6	C2	*Hi* 5'	CTCAAAATACTCTTCTGCAGTGACAAATCCATAAACTCTACCAATAAGAT	*At4g21705*
3-9	H4	*Hi* 5'	AAAGAAAACAAAAGTTAAAACGGAATATAAATATACTGTATCGAAACTAA	*At2g13290* upstream
3-10	H4	*Hi* 5'	CGAGTCCACTAAGTAGCTTCTCGCGCAAGAATCATTTTTAAAAAAATATA	*At2g23830* upstream
3-11	H4	*Hi* 5'	AGATCGATTGCATAAGAAATGGAAGAAGAAGACTGCGAAGGTTTCGAAGA	*At3g30751* downstream
3-12	H4	*Hi* 5'	AATTGGAATTATCTATCAGTCGCGTCAATTTCTGAGTGTAATAGATAGAA	*At4g04720* upstream
3-13	H4	*Hi* 5'	ACCGCCTAGGATCAGAGATCCGGGACGAGCGCTGATGTAGTACGTCCCCT	*At4g07920*
3-14	H4	*Hi* 5'	GTGTGCGACTGTTACAGTAACACTGAAGGGTTTAAAAAAAAAAAGTCAAC	*At5g11880* upstream
5-11	J1	*Hi* 5'	CCTAAAAAACCAAAAAGTGGTTACAGTGAGAAATTACACTCACCGTTTGT	*At2g02790* upstream
5-12	J1	*Hi* 5'	GACTAAAGGTGACTATGAAAAGTTGACGAATGGGAAACAGATACGGGTTG	*At3g32110*
6-1	D2	*Hi* 5'	GGTTGTAAACTATGGTTAGCTTGGATTGGTTAGGTTGGGTTAGGTTGGGT	*At5g33405*
7-1	J4	*Hi* 5'	CTCCAAAAGTCTATTCAAATGGCAAAGACAGTTGAATCCATCATCATCAT	*At5g34930* upstream
8-1	H3	*Hi* 5'	CATTCATCTCCCTAACTTTGTTAACTTTGCAGTTTATTCTTTCAAATTTT	*At5g35150*

Table 1 Whole genome re-sequencing and suppression PCR were used to find *de novo* integration sites of *Hi* and other 6 copies with high sequence similarity at *ddm1* lines.
The unique terminal sequences were used to research flanking sequence of *Hi* and AT4TE15615 (top table). The details of integration site are shown (whole genome re-sequencing, middle table; suppression PCR, bottom table). Integration number, the number indicate each integration sites at Fig 2A. Line name, the line names of *ddm1*. Read_number, how much reads were found at whole genome re-sequencing. Terminal_sequence, which unique terminal sequence (5' or 3') was used to find flanking sequence. Flanking_sequence, detail flanking sequences were shown.

Integration site, the nearest genes located to each integration sites were shown.

Hi 3'	ACAAAAAAGGGACTGGCCAGACGATTATTC
Hi 5'	CCACAAAAGGGACACAAGAGATGATCCCTC
AT4TE15615 5'	GAATAATAATCTCTTGTGTCCCTTTTTTGG
AT4TE15615 3'	CCAAAAAAGGGACACAAGAGATTATTATTC

intergration number	line name	sequence	reads	Terminal_sequence
1-1	A2	TTTGGACAGAGCCAAGGTCA	2	Hi 5'
1-2	A2	CACTGACTTGCTGATCGTAT	24	Hi 5'
	A2	AAGTCAGTGTAAGATTGATT	38	Hi 3'
1-3	A2	TGAGGAAAAGCGTGGGAAGATTAAAATGACGGATATGTCCTTATCATGGT	-	Hi 5'
	A2	TTTTCCTCAAGGGCTTATTC	7	Hi 3'
1-4	A2	CTCCTTTCACCACCAATCTCTCGTTTCATTTACTACAACTACAACACTCT	-	Hi 5'
1-5	A2	AGTTTCAGATGAGAGAATAT	1	Hi 5'
1-6	A2	TACTTAAAGTGCTACGAAGA	1	Hi 5'
2-1	C2	TTTGCAAGATCTACTAAATT	2	Hi 3'
2-2	C2	ATTTAGATGCTAGGTGATTT	2	Hi 3'
2-3	C2	TATTTAAAGCTTGTGAATTC	1	Hi 3'
2-4	C2	TGTGAGAATAATCGTCTGGC	2	Hi 5'
2-5	C2	TGTGAGGAGGAACTAAGAGT	1	Hi 5'
2-6	C2	CAAAATATTAGGGTTTAATT	2	Hi 5'
2-7	C2	GTCAAGTACGCGGTCGAGTGAAACGGTCGAGTATA	7	Hi 5'
2-8	C2	CCTTCAAACCGCCTAGGATCAGAGATCCGGGACG	7	Hi 5'
	C2	GTTTGAAGGCCTAGGTCACT	13	Hi 3'
2-9	C2	TAATATAGAACGCTTAATAT	1	Hi 5'
2-10	C2	CTACGTCGACCCGTGAATCTGTTGGCCGAAC	12	Hi 5'
2-11	C2	CTCAAAATACTCTTCTGCAGTGACAAATCCATAAACTCTACCAATAAGAT	-	Hi 5'
3-1	H4	AAAGAAAACAAAAGTTAAAA	12	Hi 5'
	H4	AAAGAAAACAAAAGTTAAAACGGAATATAAATATACTGTATCGAAACTAA	-	Hi 5'
	H4	GTTTTCTTTCGATAAAGATC	10	Hi 3'
3-2	H4	TTCGGATGGTTTGGTTCGGA	2	Hi 3'

3-3	H4	TGTGAGAATAATCGTCTGGC	5	Hi 5’
3-4	H4	GATGAGGGAGAAGAGAGATG	1	Hi 3’
3-5	H4	CCTTCCGAGTCCACTAAGTA	44	Hi 5’
	H4	CTCGGAAGGAGAAGGGAAAT	24	Hi 3’
3-6	H4	CGAGTCCACTAAGTAGCTTCTCGCGCAAGAATCATTTTTAAAAAAATATA	–	Hi 5’
3-7	H4	ACTTCAGTCACACGTTTTTC	2	Hi 3’
3-8	H4	CATACGATTAAAGAAACCCA	15	Hi 3’
	H4	AATCGTATGTATACTCACCC	12	Hi 5’
3-9	H4	AATTAAAAGAATCCCTGGAC	4	Hi 3’
3-10	H4	AGATCGATTGCATAAGAAATGGAAGAAGAAGACTGCGAAGGTTTCGAAGA	–	Hi 5’
3-11	H4	CGAATCATTCTTGAGGAGAT	3	Hi 5’
3-12	H4	AATTGGAATTATCTATCAGTCGCGTCAATTTCTGAGTGTAATAGATAGAA	–	Hi 5’
3-13	H4	ACCGCCTAGGATCAGAGATCCGGGACGAGCGCTGATGTAGTACGTCCCCT	–	Hi 5’
3-14	H4	GTGTGCGACTGTTACAGTAACACTGAAG	–	Hi 5’
3-15	H4	GTTTTCGAGTCGAATATGACTTGTTGTCA	1	Hi 5’
4-1	J1	AAACATAACATGAGTGTACG	2	Hi 3’
4-2	J1	GAGATTGGAAGAACTTGACA	1	Hi 5’
4-3	J1	CTCAAATACTCATCTGTAAG	1	Hi 5’
4-4	J1	CGGGATCCACTTTGTTATAAACCTAAGTATCTGCAATTAGGGTTGTTGCTACTTCAC	1	Hi 3’
4-5	J1	TTCATTCCGTTAGAAAAGTG	2	Hi 3’
4-6	J1	ATAATCAAGTGTTTGTGTTG	1	Hi 5’
4-7	J1	CCTAAAAAACCAAAAAGTGGTTACAGTGAGAAATTACACTCACCGTTTGT	–	Hi 5’
4-8	J1	GGCCAGACGATTATTCTCAC	1	Hi 5’
4-9	J1	TCTGGCCAGTCCCTTTTTTGTAAA	1	Hi 5’
4-10	J1	TGTGAGAATAATCGTCTGGC	4	Hi 5’
4-11	J1	GAACCGAGTTAGTAGTTACA	1	Hi 3’
4-12	J1	TGTGAGGAGGAACTAAGAGT	2	Hi 5’
4-13	J1	TAAATAACATCAACATGCAT	3	Hi 3’
4-14	J1	TTCAAAATCAGATGTTCTTG	1	Hi 5’
4-15	J1	ATTTCTCAATCGAAACAGAATGTACA	2	Hi 5’
4-16	J1	GACTAAAGGTGACTATGAAAAGTTGACGAATGGGAAACAGATACGGGTTG	–	Hi 5’
4-17	J1	CAAATATACTCTCTTTTCTA	1	Hi 5’

4-18	J1	GAGACACACGCTCTGCCATT	1	Hi 5’
4-19	J1	GTTCCTTTCCCTCTCTGAAA	1	Hi 3’
4-20	J1	GTGTGCGACTGTTACAGTAA	2	Hi 5’
4-21	J1	GTGAAGGAGGGGAAGGCAGG	2	Hi 5’
4-22	J1	CTCCAAAAGTCTATTCAAAT	21	Hi 5’
	J1	CTTTTGGAGGTTTGTGGAGA	5	Hi 3’
4-23	J1	TTCAAAAAACCCACACACAC	1	Hi 3’
4-24	J1	CTTTCTCAGTCTCTCTCTGT	1	Hi 5’
5-1	L2	GTTTTCGTAGCCATGGCTTC	1	Hi 3’
5-2	L2	GCCTTAGCAAGTGCATCGGC	2	Hi 3’
5-3	L2	GTTCTGTGAGAATAATCGTC	1	Hi 5’
5-4	L2	TGTGAGAATAATCGTCTGGC	2	Hi 5’
5-5	L2	TGTGAGGAGGAACTAAGAGT	1	Hi 5’
5-6	L2	GACTGAAGTGCTTTTTTGAT	33	Hi 5’
	L2	ACTTCAGTCACACGTTTTTC	22	Hi 3’
5-7	L2	GACTAAAGGTGACTATGAAA	58	Hi 5’
	L2	CCTTTAGTCGGACCGAACAC	35	Hi 3’
5-8	L2	GTCGGTCGACACTGTCGCGTTCTGTAGTGTCTACCGTGGCTGG	2	Hi 3’
5-9	L2	GTTTCCACTCTCTCAAGGTG	1	Hi 5’
	L2	AGTGGAAACGCTACAGAGTT	1	Hi 3’
5-10	L2	CTTATGCGACCTGTTGTTAC	2	Hi 5’
5-11	L2	CTTTTTATTTCAGAGCAAAA	16	Hi 5’
	L2	AATAAAAAGAGTAAGAGAAG	10	Hi 3’
6-1	D2	GGTTGTAAACTATGGTTAGCTTGGATTGGTTAGGTTGGGTTAGGTTGGGT	-	Hi 5’
7-1	J4	CTCCAAAAGTCTATTCAAAT	-	Hi 5’
8-1	H3	CATTCATCTCCCTAACTTTGTTAACTTTGC	-	Hi 5’
AT4TE15615-1	A2	TATAGCTCAAATCGCCTTAA	1	AT4TE15615 5’
AT4TE15615-2	C2	TGATCTTTCCCCTTCTCTGACAAGGATCA	1	AT4TE15615 5’
AT4TE15615-3	J1	GATATGAAGACTACAAAAGGATTTTACA	1	AT4TE15615 5’