Thesis for Ph.D. Degree

The characterization of a heat-activated retrotransposon *ONSEN* and the effect of zebularine in adzuki bean and soybean

「アズキとダイズにおける高温活性型レトロトランスポゾン *ONSEN* の特徴とゼブラリンの効果について」

by

Patwira Boonjing

Graduate School of Life Science
Hokkaido University

June 2020

Thesis for Ph.D. Degree

The characterization of a heat-activated retrotransposon *ONSEN* and the effect of zebularine in adzuki bean and soybean

「アズキとダイズにおける高温活性型レトロトランスポゾン *ONSEN* の特徴とゼブラリンの効果について」

by

Patwira Boonjing

Graduate School of Life Science
Hokkaido University

June 2020

Abstract

The *Ty1/copia*-like retrotransposon *ONSEN* is conserved among *Brassica* species, as well as in beans, including adzuki bean (*Vigna angularis* (Willd.) Ohwi & Ohashi) and soybean (*Glycine max* (L.) Merr.), which are the economically important crops in Japan. *ONSEN* has acquired a heat-responsive element that is recognized by plant-derived heat stress defense factors, resulting in transcribing and producing the full-length extrachromosomal DNA under conditions with elevated temperatures. DNA methylation plays an important role in regulating the activation of transposons in plants. Therefore, chemical inhibition of DNA methyltransferases has been utilized to study the effect of DNA methylation on transposon activation. To understand the effect of DNA methylation on *ONSEN* activation, *Arabidopsis thaliana*, adzuki bean, and soybean plants were treated with zebularine, which is known to be an effective chemical demethylation agent. The results showed that *ONSEN* transcription levels were upregulated in zebularine-treated plants. Extrachromosomal DNA of *ONSEN* was also accumulated in the treated plants. However, it depends on many factors such as the concentration of zebularine, the cultivation area, or the kind of plant tissue. Another factor which might involve in *ONSEN* transcription and transposition was chromosomal location. In soybean, most of the *ONSEN* copies were located in pericentromeric regions which is chromatin repressive regions.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Transposable elements (TEs)

Transposable elements (TEs) are the major component of the genome in many eukaryote species (Wessler, 1996). TEs are also divided into two classes, including class I: Retrotransposon that generally function via reverse transcription by a copy and paste mechanism and class II: DNA transposon encodes the protein transposase that is required for insertion and excision by a cut and paste mechanism. They are further subdivided into 9 orders and 29 super families as showed in Figure 1.

We are focusing on the Class I : Retrotransposon that is divided into two families depending on whether the Long Terminal Repeat (LTR) is present or absent at either end of the element (Finnegan, 1989). The long interspersed nuclear elements (LINEs) is considered to be a representative in non-LTR retrotransposons group (Goodier and Kazazian, 2008). The reverse transcriptase uses the elements' mRNA poly-A tail as a primer and starts DNA strand synthesis directly at the point of insertion and it also has the ability to ligate the end of the newly synthesized DNA into the insertion point. LTR retrotransposons has the insertion of the DNA copy via integrase enzyme, which is encoded by the element itself. At the target site, staggered cuts are made. Extra nucleotides from the 3′ termini of the LTRs are trimmed, finally, the 3′ termini are joined to the free 5′ ends at the staggered cut (Schulman, 2012) (Figure 2).

The lifecycle of LTR retrotransposon is started from reverse transcribed RNA in the nucleus. The transcripts are exported to the cytoplasm that are translated into the polyprotein encoded by the retrotransposon. The RNA template is packaged into virus-like particles (VLPs) consisting of the *gag* gene products associate with transcripts of the LTR retrotransposon to form virus-like particles while the *pol* gene encodes integrase, reverse transcriptase, and RNase H domains. The *gag* and *pol* genes overlap so that translation of *pol* requires a frameshift as is the case for retroviral transcripts. The VLP with RNA is reverse transcribe to cDNA, then, the VLP with cDNA is imported into the nucleus and the integration takes place (Figure 3). The

LTR retrotransposons are divided into two super families consisting of *Gypsy* and *Copia* which most likely originated from two independent gene fusion events.

LTR has three functional regions including U3, R, and U5 (Figure 4). The U3 region contains the promoter and located upstream from the transcription start site (TSS). In addition, it contains most of the regulatory motifs involved in transcriptional regulation. From the TSS, the U5 region is located downstream in the 3′ LTR. The remaining sequence located between U3 and U5 is called R region. Indeed, 5′ and 3′ LTR are generated from the reconstitution of the full LTR after reverse transcription, therefore, they are identical in recent copies. The 3′ LTR could release readout RNA into adjacent sequences which is able to regulate the expression of nearby genes (Schulman, 2013).

In flowering plants, the activation of TEs is under epigenetic control via small RNA-mediated epigenetic pathways or RNA-directed DNA methylation (RdDM) (Figure 5). RdDM contains specific DNA-dependent RNA polymerase, namely PolIV, which produces an initial RNA transcript for RNA silencing and PolV that transcripts a messenger RNA for DNA methylation on the target site. RNA-Dependent RNA Polymerase 2 (RDR2) produces double stranded RNA and Dicer-Like 3 (DCL3) processes to 24-26nt small-interfering RNAs (siRNAs). The siRNAs bind to Argonaute 4 (AGO4) that interact with polV to recruit Domains Rearranged Methyltransferase 2 (DRM2) leading to de novo DNA methylation of the target TEs (Matzke & Mosher, 2014).

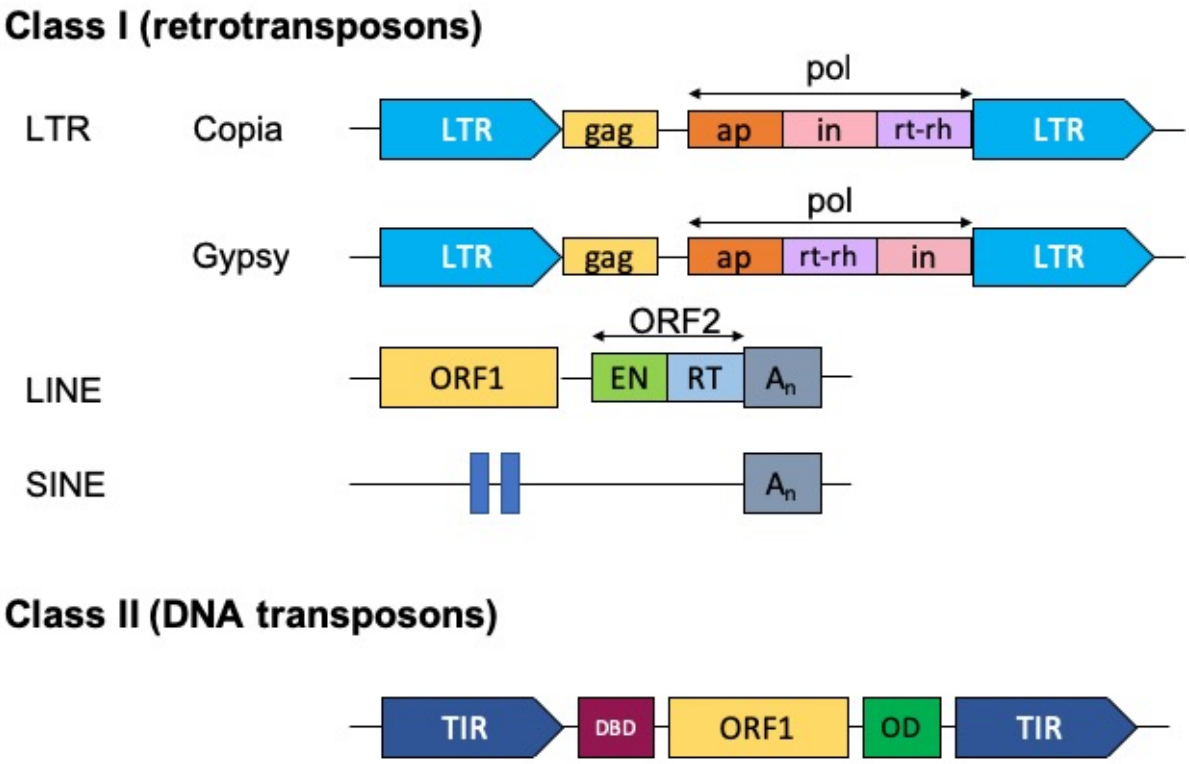

Figure 1 Classification system for TEs (modified from Wicker et al., 2007). LTR: long terminal repeat, gag: group-specific antigen, pol: polyprotein, ap: aspartic protease, in: integrase, rt-rh: reverse transcriptase – RNaseH, ORF: open reading frame, EN: endonuclease, RT: reverse transcriptase, A_n: a poly(A) tail sequence, TIR: terminal inverted repeat, DBD: DNA-binding domain. The blue boxes within the SINE retrotransposon indicate tRNA related regions.

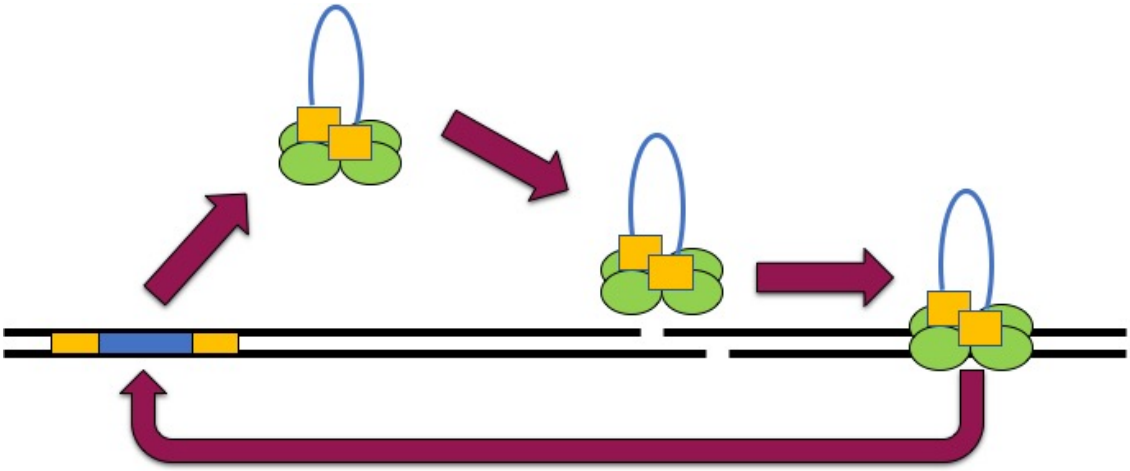

Figure 2 Integration mechanism of an LTR retrotransposon (modified from Schulman, 2012). The yellow box, blue box, and green circles indicate LTR, ORF, and integrase regions, respectively. The retrotransposon is represented as a loop bounded by two LTRs. Each LTR is flanked by an extra dinucleotide basepair, which is copied by RT from the dinucleotide found between the PBS and the 3′end of the 5′LTR during reverse transcription. The integrase is represented, bound to the LTRs, as a tetramer, forming a pre-integration complex together with the retrotransposon. The genomic DNA target is shown as a blue line beneath the preintegration complex. The integrase makes a 4-to 6-bp staggered cut in the genomic DNA and trims the dinucleotide from the 3′end of each LTR, generating 5′overhangs on both the retrotransposon and at the target site.

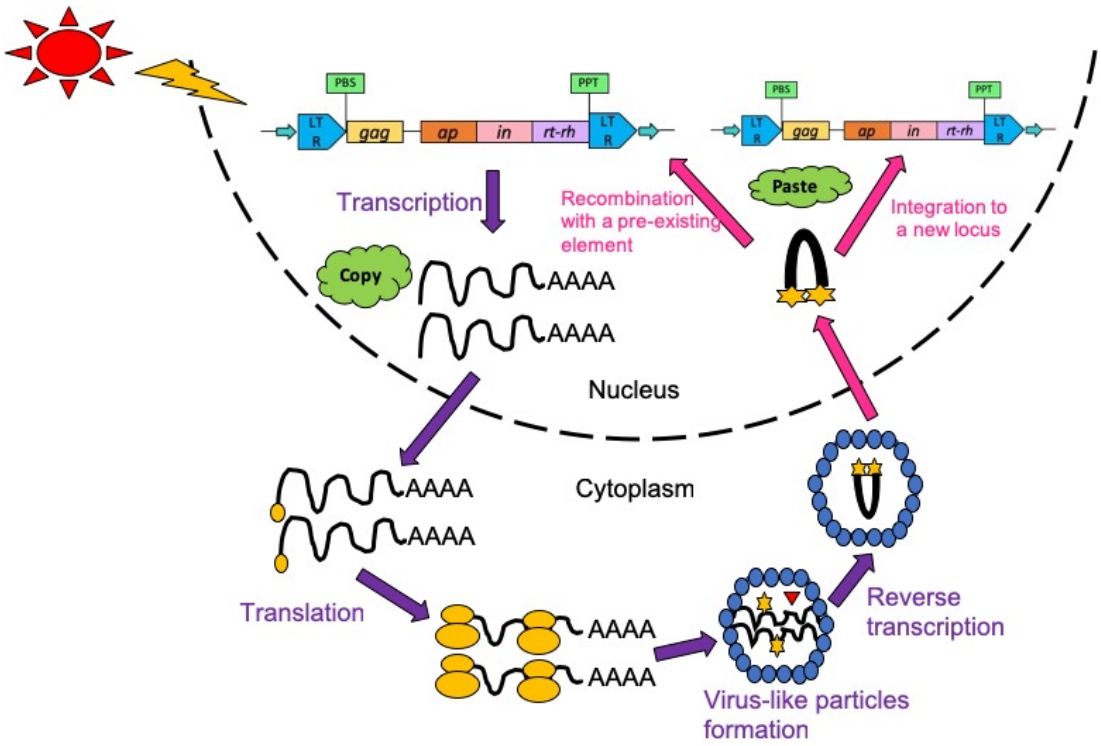

Figure 3 Lifecycle of LTR retrotransposons (modified from Saariaho, 2006).

The integrated retroelement is transcribed into mRNA (black line) and exported to the cytoplasm, where it is translated into GAG and POL polyproteins that are processed into functional units by an element-encoded protease (PR). These units and cellular tRNA, which acts as a primer for reverse transcription, are assembled into virus-like particles (VLPs) together with the transcript (mRNA depicted by black curved line) that is then converted to cDNA by reverse transcriptase (RT) within the VLPs. This cDNA is finally transferred into the nucleus in the context of a preintegration complex that also contains integrase (IN) that finally integrates the cDNA as a new copy into the host genome or recombines with a pre-existing element.

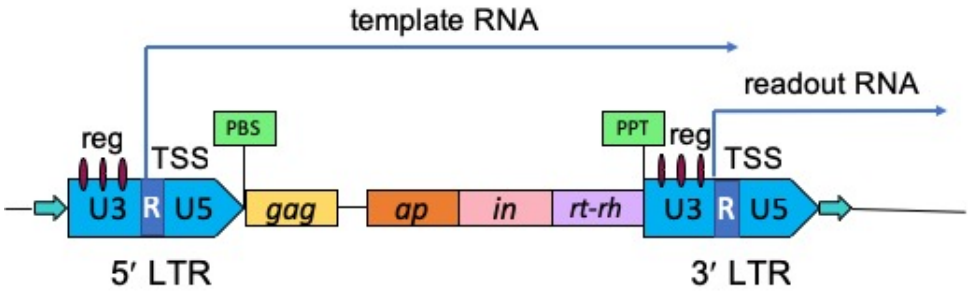

Figure 4 Structure of LTR retrotransposon (modified from Schulman, 2013). LTRs are generally composed of U3, R, and U5 domains, each one with a specific function in the retrotranscription process. R and U5 sections are generally more conservative than U3, probably due to the adaptation to varying tissue environments and to different stress responses. RNA template is generated from R to R sections, it contains only one U5 and U3 section, and eventually, two identical LTRs when the DNA copy of the element is inserted into the genome. LTR: long terminal repeat, reg: regulatory motifs, TSS: transcription start site, PBS: primer binding site, PPT: polypurine track, gag: group-specific antigen, pol: polyprotein, ap: aspartic protease, in: integrase, rt-rh: reverse transcriptase – RNaseH.

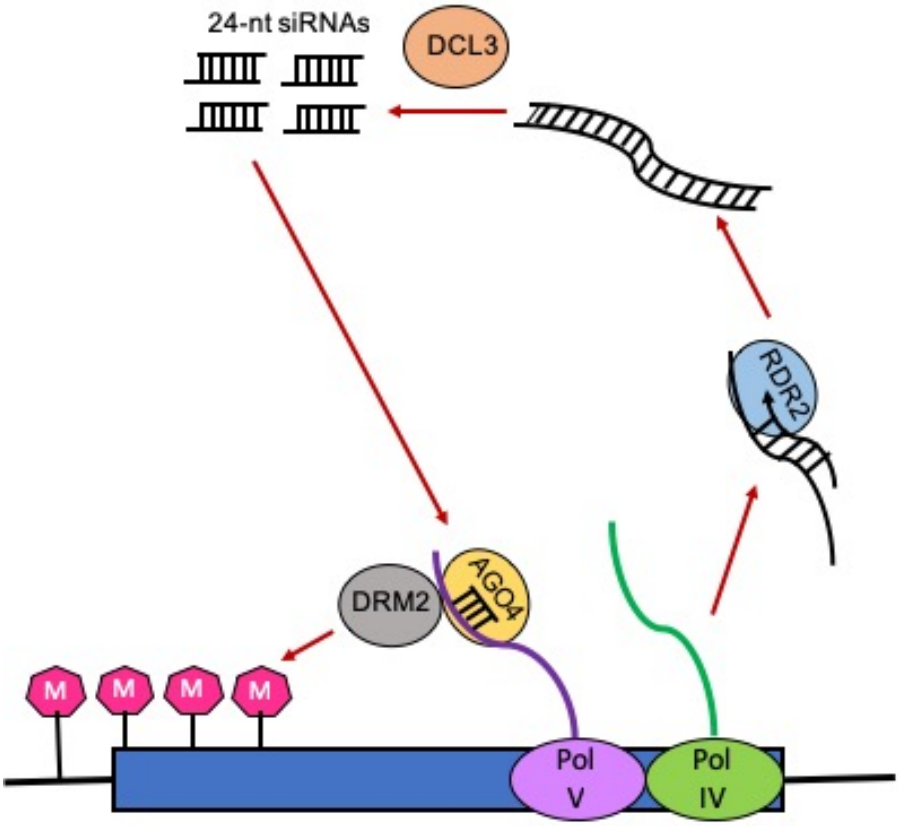

Figure 5 RdDM pathway mechanism (modified from Matzke & Mosher, 2014). RNA transcripts are generated from repetitive sequences (transposons and others) by an RNA polymerase known as Pol IV. RNA-Dependent RNA Polymerase 2 (RDR2) then converts the RNA to double stranded transcripts. These are processed into 24-nucleotide small RNAs (siRNAs) by Dicer-Like3 (DCL3). These are methylated at their 3 ends and the guide strand complementary to the genomic DNA, that will be the target of the RdDM, is incorporated into Argonaute 4 (AGO4). AGO4 is recruited through interactions with Pol V. AGO4 and Domains Rearranged Methyltransferase 2 (DRM2) are connected, then de novo methylation of DNA is catalyzed.

1.2 Stress responsible transposable elements in plants

Many LTR retrotransposons in plants response to various stimuli such as tissue culture or external stress either biotic or abiotic. The environmental stresses influence plant growth and development by changing the gene expression, which can include transposon regulation. Both up and downregulation of TEs could be detected depending on the type of the stress. The most popular environmental stresses consist of salt, cold, and heat stress. For example, in *Arabidopsis thaliana*, *AtCopeg1* elements led to upregulation by various nutrition starvation, salt stress, and cytokinin, however, the expression was downregulated by abscisic acid (Duan et al., 2008).

Salt stress-tolerant plants have evolved different adaptive mechanisms to display different degrees of tolerance. For example, one of the *Ty1/copia* retrotransposon, namely, *Ttd1a* transcript in leaves of *Triticum durum* showed the response both activation and mobilization to salt stress because salt stress-induced transcription factor could bind to a promoter of *Ttd1a* retrotransposon in durum wheat. (Woodrow et al., 2010)

Cold stress was reported to down regulate DNA methyltransferase (*ZmMET1*) in root tissues resulting in hypomethylation in an Ac/Ds transposon. Locus-specific siRNA-mediated RdDM was triggered by the transcriptional activation from the transposons (Marocco et al., 2005). Another example was shown in the *Tam3* elements in *Antirrhinum majus* that was activated at low growth temperatures of 15°C, while, the activity was suppressed at high growth temperature of 25°C. Consequently, the methylation level at 15°C was lower than at 25°C. It was indicated that siRNA-mediated methylation might decrease, re-expression of silenced transposons might occur at low temperature (Coen et al., 1989).

Some LTR retrotransposons are silenced under normal condition and activated by tissue culture such as callus induction (Hirochika et al., 1996). *BARE* in barley and *Tos17* in rice are demonstrated as examples for this case. Sixty-day-old callus lead to different retrotransposons

movement (Fras et al., 2007). Plant tissue culture induces variation such as wounding, 2,4-D treatment, and synthetic medium may be responses to stress conditions (Granbastian, 1998, LoSchiavo et al., 1989).

I focus on heat stress responsible transposable elements. From the previous studies, heat stress results in increasing genetic instability and somatic homologous recombination (Pecinka et al., 2009). In addition, heat stress could affect the genetic stability through chromatin modification and the accessibility of DNA for repair and recombination (Kirik et al., 2006). Normally, transcriptional gene silencing accompanying with high levels of DNA methylation, inactive chromatin marks, and chromatin compaction suppress TE's expression (Soppe et al., 2002).

1.3 *ONSEN*

A *Ty1/copia*-like retrotransposon, namely, *ONSEN* (*AtCOPIA78*) is firstly found to be activated by heat stress in *Arabidopsis thaliana* (Ito et al., 2011). Some LTR retrotransposons consist of *cis*-regulatory sequences in their 5′-LTR that could be recognized by a stress-induced transcription factor, which triggers the expression of transposon in response to a particular stimulus (Casacuberta and Gonzalez, 2013). *ONSEN* also has a heat-responsive element recognized by plant-derived heat induced transcription factor, namely, *HSFA2* (Cavrak et al., 2014). In the promoter of *ONSEN* LTR, *HSFA2* binds to the *cis*-regulatory heat-responsive element resulting in transcription and production of full-length extrachromosomal DNA under elevated temperature (Ito et al., 2011). *ONSEN* is not only activated in *Arabidopsis* but also in other plants such as in Brassicaceae family (Ito et al., 2013) and adzuki bean (Masuta et al, 2018).

In Brassicaceae, *ONSEN* sequences were conserved in 10 of 12 species including *Arabidopsis lyrata*, *Arabidopsis halleri*, *Crucihimalaya wallichii*, *Arabidopsis korshinsky*, *Turritis glabra*, *Arabis hirsute*, *Lepidium sativum*, *Sisymbrium irio*, *Thlaspi arvense*, and *Raphanus sativus*. In addition, a conserved heat responsive *cis*-element, *HSFA2*, in each species was also detected (Ito et al., 2013). Transgenerational transposition of *ONSEN* has been reported to occur in higher rate in regenerated plant from callus than in seedling after subjecting to heat stress (Ito et al., 2011), therefore, transposition might be regulated by multiple factors involved in gene expression and epigenetic modification during plant dedifferentiation and redifferentiation (Neelakandan and Wang, 2012). The regulation of *ONSEN* might be altered by the RdDM-mediated pathway in callus because the expression of RdDM-related genes was not suppressed and some of them were upregulated in callus. Moreover, the amount of heat-activated *ONSEN*-derived small RNAs in callus also supported this hypothesis. It showed 10-fold higher than in seedlings (Masuta et al., 2017).

In legume crops, *Ty1/copia*-like retrotransposons have also been found. Genome sequences for pigeon pea (*Cajanus cajan*) has been reported and the LTR retrotransposon named *Panzee* showed LTR which flanks its internal region (Lall & Upadhyaya, 2002). In addition, two *Ty1/copia* retrotransposon families, *pvRetro3* and *pvRetro4*, were found to have an extra ORF located between the transposase ORF and 3′LTR in common bean (*Phaseolus vulgaris*). Another example of *Ty1/copia* retrotransposon is *Soybean RetroElement 1* (*SORE-1*) which was found in soybean. This element was first discovered in an exon of the GmphyA2 gene encoding phytochrome A and conferring photoperiod sensitivity (Liu et al., 2008). Adzuki bean (*Vigna angularis* (Willd.) Ohwi & Ohashi) is one of the economic crops in Japan, especially in Hokkaido. To enhance the genetic diversity of this bean, mutagenesis by using TEs might be useful. Since full-length genome sequencing is available in adzuki bean, it is suitable to find transposable elements (Sakai et al., 2015)(Figure 6). Moreover, it has been reported that heat activated *ONSEN* in adzuki bean is expressed in two varieties consisting of Shumari and Tanba Dainagon although extrachromosomal DNA (ecDNA) was detected only in the Tanba Dainagon variety (Masuta et al., 2018).

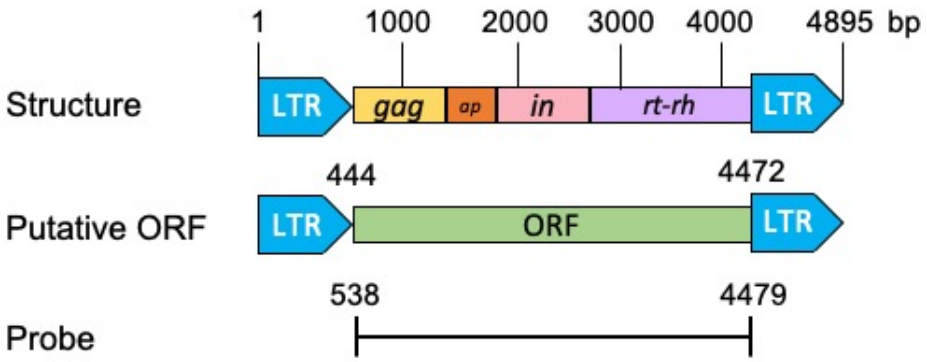

Figure 6 The schematic figure of full-length *ONSEN*-like elements in *V. angularis* (modified from Masuta et al., 2018). The schematic figure of full-length *ONSEN*-like elements in *V. angularis*. GAG: gag polyprotein, PR: protease, INT: integrase, RT: reverse transcriptase, RH: ribonuclease H. probe: A Southern probe designed in the open reading frame (ORF).

1.4 DNA methylation

Epigenetic modification including DNA methylation and histone modification are usually repressed the activities of TEs (Chandler and Walbot, 1986). In plant genomes, DNA methylation has been reported to be enriched in TEs (Rabinowicz et al., 2003). DNA methylation occurs in all sequence contexts (CG, CHG, CHH, where H is A, T, or C). The function has been mainly implicated in silencing of repetitive DNA (Feng et al., 2010). Several DNA methyltransferases showed preference for cytosine methylation in particular DNA sequence contexts and acted via different mechanisms (Stroud et al., 2013).

1.5 The chemical demethylating agents

There are various kinds of drugs that interfere with chromatin marks (Figure 7) such as methyl-group biosynthesis inhibitors, DNA methylation inhibitor, histone deacetylase inhibitors and RNA polymerase II inhibitors. However, I focus on DNA methylation inhibitors because plant DNA methylation is transmitted throughout development (Feng et al., 2010). In addition, some TEs are reported to recruit DNA methylation via RdDM pathway **(**Matzke & Mosher, 2014). TEs are activated after subjecting to heat stress or DNA methylation inhibitors, leading to higher level of gene transcription level (Chang and Pikaard et al., 2005) and extrachromosomal DNA copy number (Masuta et al., 2018).

The most widely used chemical demethylating agents to inhibit DNA cytosine methylation and reactivate silenced genes consist of 5-azacytidine (5-AZA), 5-aza-2′-deoxycytidine (5-AZA-2′) (Jones and Taylor, 1980), and zebularine (ZEB) (Driscoll et al., 1991) that act as non-methylable cytosine analogs, incorporating into the DNA double helix in the place of cytosine with each cycle of DNA replication.

Both 5-AZA and 5-AZA-2′ have a nitrogen in place of a carbon at position 5 of the pyrimidine ring (Figure 7). Nevertheless, 5-AZA has a short half-life of only 4 hours. Therefore,

it needs to be supplied several times in typical experiment (Champion et al., 2010). In addition, 5-AZA is unstable in neutral aqueous solutions and the hydrolysis products have been well characterized (Beisler, 1978). Other analogs have been developed, for example, 5-Fluoro-2′-deoxycytidine is able to inhibit DNA methylation and reactivates silenced genes when incorporated into DNA (Jones and Taylor, 1980). However, generating 5-fluorodeoxyuridine and its metabolites may be toxic (Boothman et al., 1989).

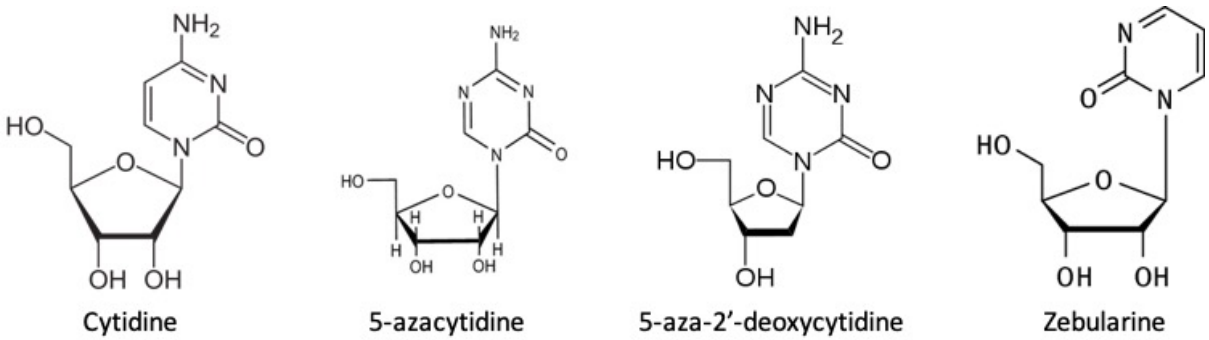

Figure 7 Structure of cytidine and its analogs

1.6 Zebularine

One of the DNA methylation inhibitor, zebularine (ZEB) is a cytidine analog consisting of a 2-(1-H)-pyrimidinone ring (Figure 7) (Driscoll et al., 1991). ZEB's half-life is over 3 weeks (Champion et al., 2010) and is stable in up to pH12 aqueous solution (Barchi et al., 1992). ZEB was also found to form tight complexes with bacterial methyltransferases, resulting in a strong DNA methylation inhibition (Hurd et al., 1999). Therefore, it is a suitable inhibitor in plant studies. In *A. thaliana*, *TS-GUS* line containing a transcriptionally silent ß-glucuronidase transgene that is activated throughout the entire plant in the epigenetic mutants background was used to apply to 20 or 40 μM zebularine containing medium, resulting in GUS silencing in cotyledons but neither in true leaves of all stages nor in floral tissues. In addition, zebularine treated plants indicated that zebularine treatment reduced DNA methylation of centromeric repeats in both tissues but heterochromatin condensation is still higher in true leaves (Baubec et al., 2014). Moreover, another study showed that zebularine treatment lead to concentration-dependent loss of DNA methylation genome-wide in all context (CG, CHG, CHH). DNA methylation was also reduced across chromosome1, especially in pericentromeric region when zebularine was treated (Griffin et al., 2016).

1.7 Chromosomal distribution in soybean

LTR retrotransposon activation occurred by random insertion (Masuta et al., 2017) and by insertion into gene-rich regions and pericentromeric regions (Paterson et al., 2009), however, the evidence for preferential insertion in these regions is still limited (Nakashima et al., 2018). In addition, factors which affected the activation could be different depending on LTR retrotransposon family (Pereira, 2004), age of the elements (Xu and Du, 2014), or copy number (Baucom et al., 2009).

Recently, a genome-wide survey of transposable elements has been done in soybean (*Glycine max* L.) cultivar Williams 82 (Wm82) which is in the similar family to adzuki bean (Fabaceae). Since whole genome database of adzuki bean has not been available, using soybean database could help to predict the *ONSEN* insertion characterization in adzuki bean. Because soybean has tolerated whole genome duplication for two rounds (Shoemaker et al., 2006), therefore, polyploidy and genome evolution studying are appropriate (Schmutz et al., 2010). In addition, the distribution of 510 LTR retrotransposon families including 32,370 elements has been analyzed (Du et al., 2010). Among 32,370 elements, 31,858 of them were anchored to the currently assembled 20 chromosome pseudomolecules (Schmutz et al., 2010). Most of these elements were present in pericentromeric regions (Tian et al., 2012).

Among LTR retrotransposon elements, there are major two types of the elements including intact elements that are complete elements and solo LTRs that used to be formed of intact elements over evolutionary time, when time passed, ORF regions are deleted remaining only single LTR region (Du et al., 2010) (Figure 8). Solo LTRs' formation is occurred by unequal intra-element homologous recombination. It is thought to be a major process for removing LTR retrotransposons in plants (Devos et al., 2002) (Figure 8). Unequal and illegitimate recombination have been thought to be common mechanisms responsible for retrotransposon rapid elimination during plant genomes evolution. Indeed, illegitimate

recombination that generates small deletions has been known to be a major mechanism for elimination of LTR retrotransposons in *Arabidopsis* (Devos et al., 2002).

Since *ONSEN* is one of *Ty1/copia* retrotransposons and there is no previous study about *ONSEN* in soybean, leading to my objective of soybean chromosomal distribution study that is to understand overall features of *ONSEN* in the soybean genome. There is a study about a *Ty1/copia* retrotransposons in soybean, named, *Soybean RetroElement 1* (*SORE-1*) that was discovered in the *GmphyA2* gene's exon encoding phytochrome A and conferring photoperiod sensitivity (Liu et al., 2008). This element was also detected in an intron of the *GmFT2a* gene which is an ortholog of *A. thaliana FLOWERING LOCUS T* (Zhao et al., 2016). The insertion time of *SORE-1* into the genome was most actively transposing at the present time indicated that *SORE-1* might be useful as an insertional mutagenesis source that could be applying for breeding. For example, it can be used to check unusual phenotype gene which may have been inactivated (Nakashima et al., 2018). Using the similar study pattern to *ONSEN* might be effective in breeding also. Other than breeding, retrotransposon insertion might interfere gene duplication. Some of gene functions may be loss, contribute to adaptive evolution (Kanazawa et al., 2009). Transposition of *ONSEN* could lead to functional diversification of various genes in the soybean genome, therefore, study about chromosomal distribution is important to predict the inserted position to find the gene function disruption which is a cause of adaptive evolution in the future.

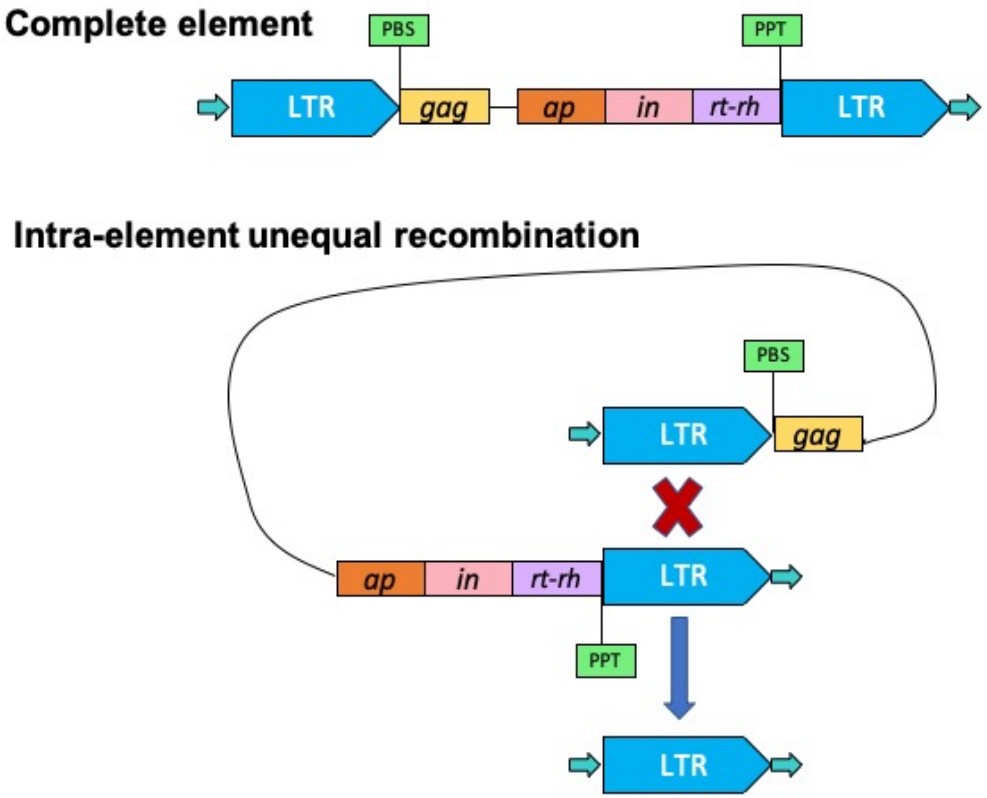

Figure 8 Unequal intrastrand recombination between LTR retrotransposons (Devos et al., 2002). (Top) Structure of a complete element, with a direct repeat (DR) of flanking target-site DNA, two long terminal repeats (LTRs), a primer-binding site (PBS), and polypurine tract (PPT) needed for element replication and encoded gene products (gag, pol). (Bottom) Solo LTR resulting from intra-element recombination. The dotted line is presented to facilitate depiction of the folding needed to accomplish this recombination and does not represent any significant stretch of DNA.

2 Results

2.1 Detection of new *ONSEN* insertion in *A. thaliana* callus treated with zebularine subjected to heat stress

To detect the new insertions of *ONSEN* that occurred after plants were treated with zebularine and subjected to heat stress, we conducted a Southern blot analysis using DNA extracted from the callus of *A. thaliana*. Many new bands were detected in 0.01 and 0.04 mM of zebularine treated samples although new bands were not detected in the control samples (0 mM of zebularine) (Figure 9A). To analyze the copy number of *ONSEN* in both 0.01 and 0.04 mM zebularine treated samples, the relative copy number of *ONSEN* was analyzed by qPCR analysis. The copy number of *ONSEN* in the 0.01 mM zebularine-treated seedlings are 7.97 folds higher than that in the non-treated seedlings subjected to heat stress. Whereas, there was no significant difference between 0.01 and 0.04 mM zebularine-treated samples subjected to heat stress (Figure 9B). The result suggested that 0.01mM zebularine enhanced the activation of *ONSEN* in *A. thaliana* callus.

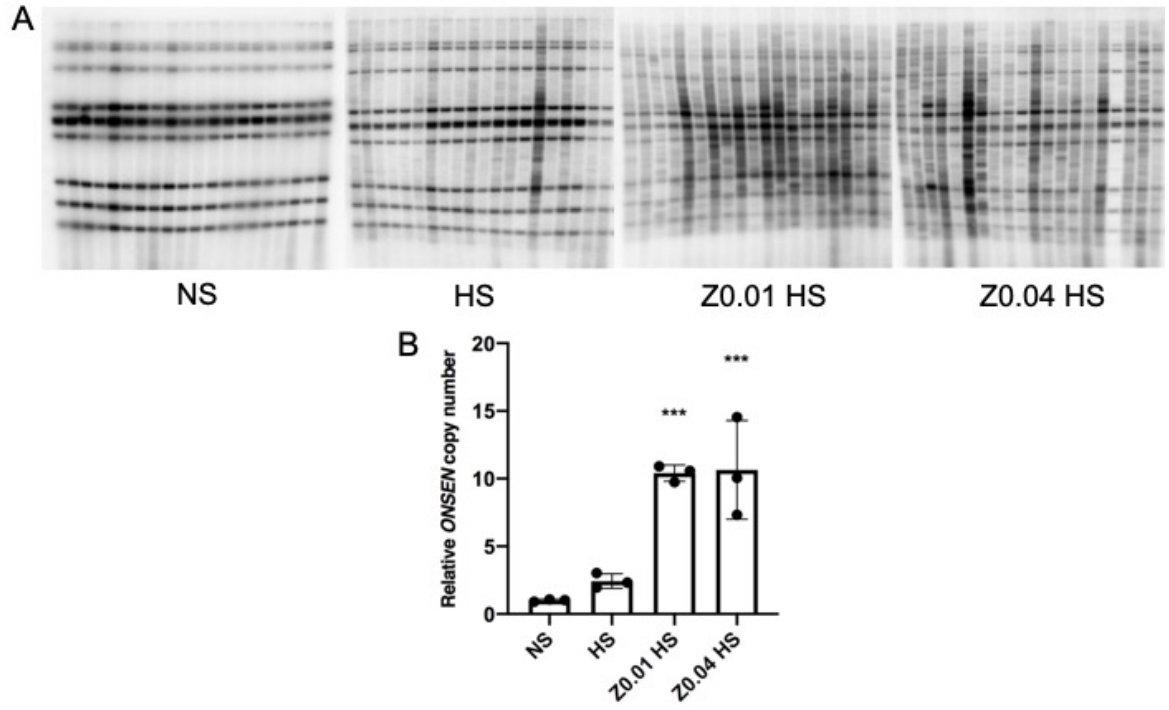

Figure 9 Effect of zebularine on *ONSEN.*

A Southern blotting of *ONSEN* copies. The DNA was digested with *Eco*RV to detect new insertion of *ONSEN* in *Arabidopsis thaliana.* The DNA was extracted from callus 160 days after heat stress under different concentration of zebularine. NS: non-stress, HS: heat stress, Z0.01 HS: heat stress with 0.01 mM zebularine treatment, Z0.04 HS: heat stress with 0.04 mM zebularine treatment. **B** The relative copy number of *ONSEN* in the callus was analyzed by quantitative PCR (qPCR). Error bars indicate standard deviations (n=3). Ordinary one-way ANOVA for qPCR data were performed. Asterisks indicate significant differences among treatments comparing with non-stress treatment: *, **, ***, and **** indicate p 0.0332, 0.0021, 0.0002, and < 0.0001, respectively (Dunnett's test). NS: non-stress, HS: heat stress, Z0.01 HS: heat stress with 0.01 mM zebularine treatment, Z0.04 HS: heat stress with 0.04 mM zebularine treatment.

2.2 The effect of zebularine on *ONSEN* expression in adzuki bean

To analyze whether zebularine affected to *ONSEN* expression, we treated a variety of adzuki bean such as Tanba-dainagon and Shumari with various concentrations of zebularine (0, 0.1, 0.3, 0.5, and 1.0 mM). In the medium containing 0.5 mM zebularine, the growth of 1-week-old Tanba-dainagon showed dwarfism in stem elongation after treating with zebularine (Figure 10). The transcript level of *ONSEN* was analyzed by qRT-PCR. The results showed that *ONSEN* expression was not activated without heat stress treatment. In Tanba-dainagon, the transcript level was gradually increased from 0 mM to 0.5 mM zebularine, but the expression level of 1.0 mM zebularine decreased to the same level as 0.3 mM (Figure 11A). While Shumari showed the nearly similar expression levels which were already high from 0.1 to 0.3 mM and the expression level also decreased after treated with 0.5 mM zebularine (Figure 11B). The results suggested that 0.3 mM zebularine was enough to activate *ONSEN* in Tanba-dainagon, however, in Shumari, the highest expression level at 0.5mM zebularine was about 60 folds comparing to non-stress condition of each concentration in both two varieties.

Figure 10 Comparison of the growth of 1-week-old adzuki bean seedlings in Tanba-Dainagon which was indicated by 0 mM (left) and 0.5 mM (right) zebularine containing medium

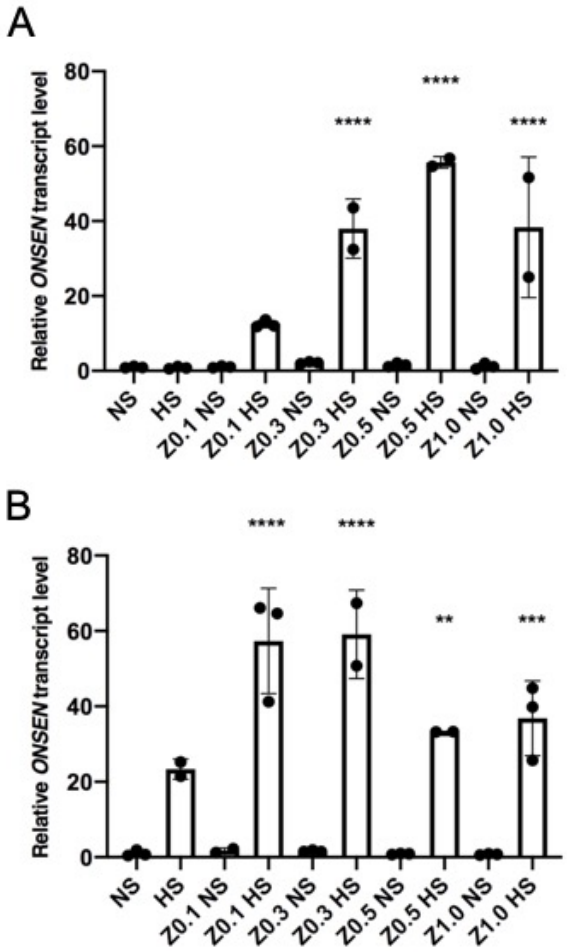

Figure 11 The relative transcript level of *ONSEN* in adzuki bean.

A Tanba-dainagon **B** Shumari after treated with 0, 0.1, 0.3, 0.5, and 1.0 mM in non-stress and heat-stress condition. NS: non-stress, HS: heat stress, Z0.1 NS: non-stress with 0.1 mM zebularine treatment, Z0.1 HS: heat stress with 0.1 mM zebularine treatment, Z0.3 NS: non-stress with 0.3 mM zebularine treatment, Z0.3 HS: heat stress with 0.3 mM zebularine treatment, Z0.5 NS: non-stress with 0.5 mM zebularine treatment, Z0.5 HS: heat stress with 0.5 mM zebularine treatment, Z1.0 NS: non-stress with 1.0 mM zebularine treatment, Z1.0 HS: heat stress with 1.0 mM zebularine treatment. Error bars indicate standard deviations (n=2-3). Two-way ANOVA for qPCR data was performed. Asterisks indicate significant differences among treatments comparing with non-stress treatment: *, **, ***, and **** indicate p 0.0332, 0.0021, 0.0002, and < 0.0001, respectively (Tukey's test).

2.3 The effect of zebularine on *ONSEN* transposition in adzuki bean

To test the transposition activity of *ONSEN* in adzuki bean plants treated with zebularine, we analyzed the extrachromosomal DNA of *ONSEN* by Southern blot analysis. Extrachromosomal DNA was detected in Tanba-dainagon subjected to heat stress. The strongest signal was detected in those treated with 0.5 mM of zebularine (Figure 12). This result agreed with that of the qRT-PCR determination of *ONSEN* transcript levels. This result indicated that 0.5 mM of zebularine induced *ONSEN* activation efficiently.

To detect the new insertions of *ONSEN* after treated with zebularine subjected to heat stress, we analyzed the copy number of *ONSEN* using the DNA extracted from the callus of Tanba-dainagon. The copy number of *ONSEN* in the 0.5 mM zebularine-treated callus are three times higher than that in the non-treated callus subjected to heat stress (Figure 13). The result suggested that 0.5mM zebularine could enhance the activation of *ONSEN* in adzuki beans.

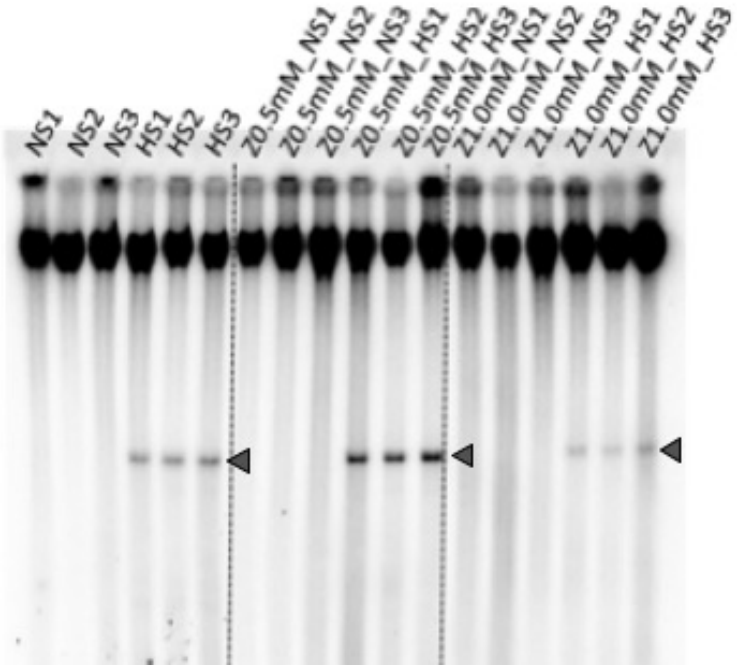

Figure 12 Effect of zebularine on the *ONSEN* activation.

The non-digested DNA was loaded to detect extrachromosomal DNA of *ONSEN* in 1-week-old adzuki seedling (Tanba-dainagon) in 0, 0.5, 1.0 mM zebularine. NS: non-stress, HS: heat stress, Z0.5mM_NS: non-stress with 0.5 mM zebularine treatment, Z0.5mM_HS: heat stress with 0.5 mM zebularine treatment, Z1.0mM_NS: non-stress with 1.0 mM zebularine treatment, Z1.0_HS: heat stress with 1.0 mM zebularine treatment. Arrows indicated extrachromosomal DNA of *VaONS*.

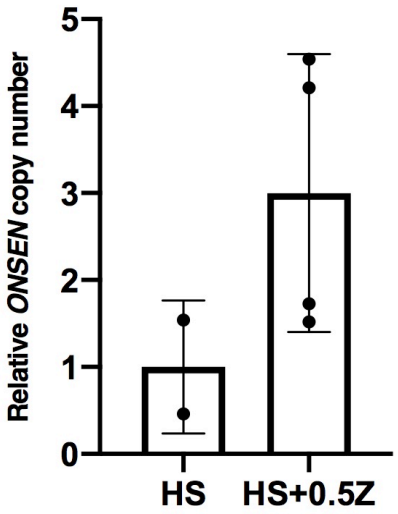

Figure 13 The relative copy number of *ONSEN* in adzuki bean callus (Tanba-dainagon) after treated with 0 and 0.5 mM in heat-stress condition which was analyzed by quantitative PCR (qPCR). HS: heat stress, HS+0.5Z: heat stress with 0.5mM zebularine treatment. Error bars indicate standard deviations (n=2-4). Unpaired t-test for qPCR data were performed.

2.4 Chromosomal location, structure, and insertion time of *GmCOPIA78*

Eighty-seven LTR elements of *ONSEN* in the Williams 82 genome of soybean were analyzed by using Soybase (http://soybase.org) which is the database of soybean (Grant et al., 2010). First of all, we performed BLAST TE Database analysis by choosing the Transposable Elements from Genome Browser Tab and also selected a database to be a Williams 82 Transposable Elements. The blast highest score elements were selected to search about the details of those transposable elements. Finally, the start and end position were copied to see the region on the Genomic Map. Chromosomal location of each element was identified. Based on the Soybase, we found that two *GmCOPIA78* families including Gmr8 and Gmr273 in the Wm82 genome (Figure 14). For solo LTRs identification, the LTR *ONSEN* intact elements in the Williams 82 genome of soybean were used to identify the solo LTRs by using BLAT (BLAST-like Alignment Tool) of Soybase. Forty-six solo LTRs, which lacked 5′ or 3′ LTR sequence were also identified. The percentage of *GmCOPIA78* containing in pericentromeric and chromosome arms region were 76.7% and 20.3%, respectively. In the Gmr8 family, 82.29% were located in pericentromeric region and 17.71% of those were in chromosome arms. In Gmr273, 72.97% and 27.03% were founded in pericentromeric and in chromosome arms, respectively (Table 1). The results demonstrated that most of the *GmCOPIA78* elements were found in pericentromeric regions than in chromosome arms.

The insertion time into the genome of each element was estimated on the basis of sequence different between 5′ and 3′ LTRs of the intact elements (SanMiguel et al., 1998), the nucleotide sequences of both of them were aligned by using the ClustalW program (Thompson et al., 1994) in MEGA7 (Kumar et al., 2016). The insertion time (T) was estimated by using the formula T=k/2r that k is the distance between LTRs estimated using the Kimura's two parameter method (Kimura, 1980). r is the mutation rate that is 1.3×10^{-8} substitutions per site per year (Ma and Bennetzen, 2004), reflecting the time after insertion into the genome.

Excluding the 46 copies of solo LTRs, only one element, Gm02 8#3, among 87 copies of *GmCOPIA78* in the Wm82 genome cannot be estimated the insertion time in pericentromeric regions due to the imperfection of 5′ or 3′ LTRs (Figure 15). The elements in Gmr273 family showed earlier insertion time than those in Gmr8 LTR-RT family, which was indicated that the elements in Gmr8 family were inserted into the genome more recently than in Gmr273. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). Bootstrap values were calculated with 500 replicates. The age of each element did not correlate with the phylogenetic evolution in Gmr8 family, however, it still correlated well in Gmr273 (Figure 15).

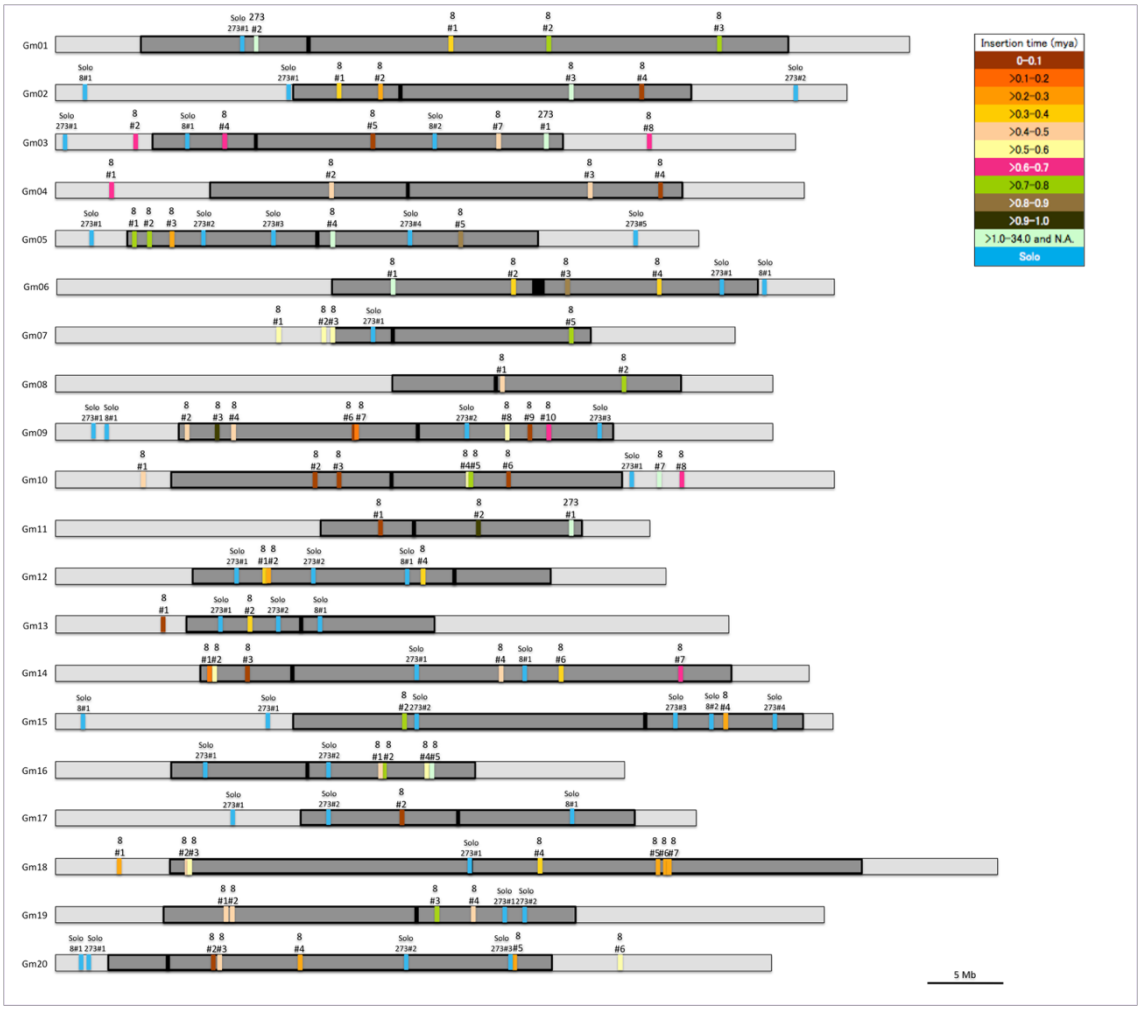

Figure 14 Chromosomal location of *GmCOPIA78* elements and solo LTRs of *GmCOPIA78*. The insertion time is indicated as colored vertical bars (dark red (the newest) to light green (the oldest)). Light blue vertical bars show the solo LTRs location. Chromosome arms and pericentromeric region are shaded in light gray and dark gray, respectively. One N.A. (not analyzed) data is not included in this figure due to the incompletion of 5′ or 3′ LTR. The number of Gm indicated chromosome number. The text over each colored vertical bar demonstrated Gmr family (the first line) and copy number of those family in each chromosome (the second line).

Family Name	Prexisting name	Super family	No. of solo LTRs			No. of intact elements			Subtotal	Ratio (S/I)	Ave. age (mys)			% of elements in pericentromeric			% of elements in chromosome arms		
			Pericentromeric	Chromosome arms	Total	Pericentromeric	Chromosome arms	Total			Pericentromeric	Chromosome arms	Total	solo	intact	total	solo	intact	total
Gmr8	GmCopia11	Copia	7	5	12	72	12	84	96	0.14	1.45	0.89	1.29	7.29	75	82.29	5.21	12.5	17.71
Gmr273		Copia	24	10	34	3	0	3	37	11.33	13.82	N/A	13.82	64.87	8.11	72.97	27.03	0	27.03

Table 1 Summary of the data of Gmr family which were found in *GmCOPIA78* sequence in Wm82 soybean. Preexisting name was the name of families denoted as numbers derived from Wawrzynski et al. (2008). Subtotal was copy numbers of intact elements and solo LTRs. Ratio (S/I) was ratio of solo LTRs (S) to intact elements (I). Ave. age (mys) was average age of the elements (million years ago).

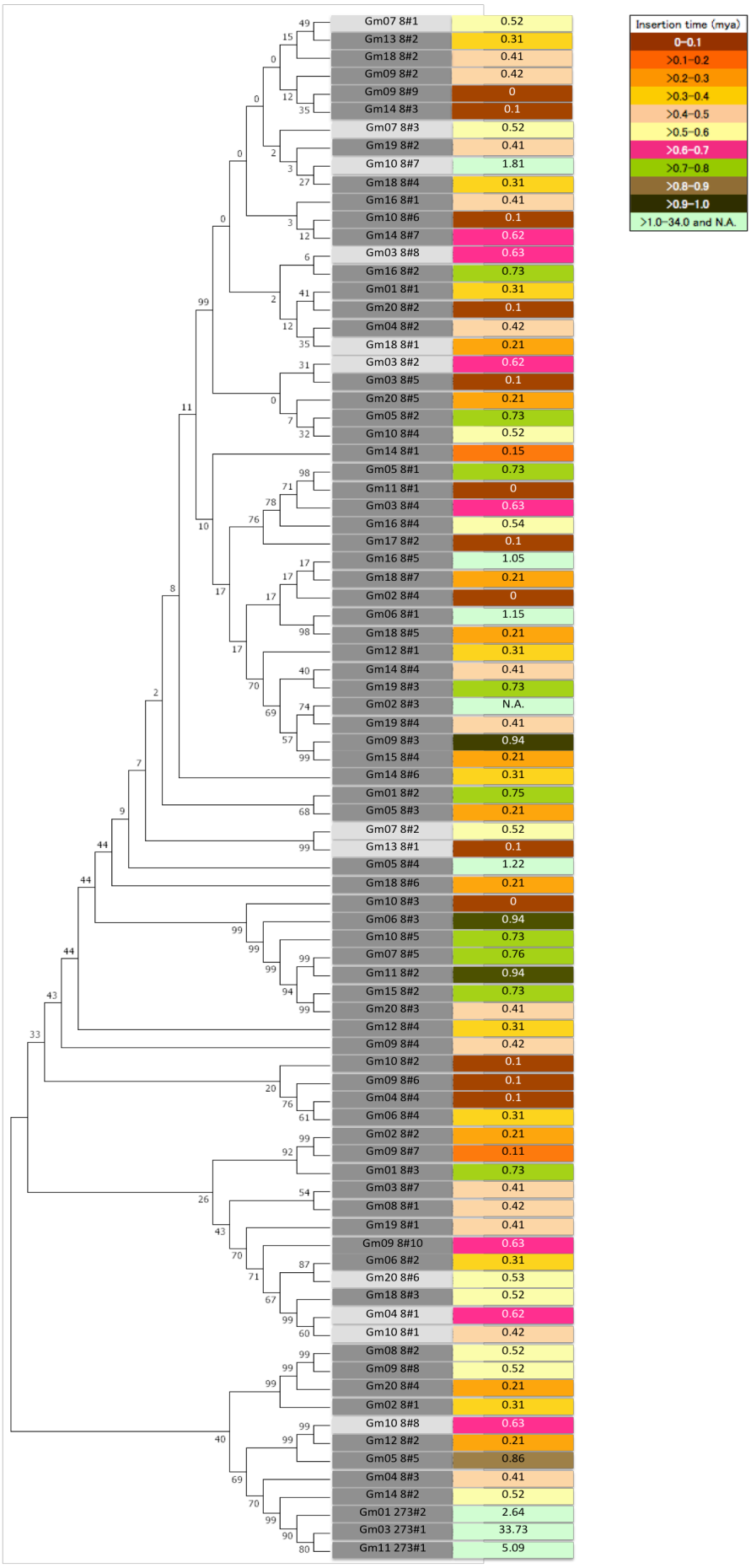
Insertion time (mya)
0–0.1
>0.1–0.2
>0.2–0.3
>0.3–0.4
>0.4–0.5
>0.5–0.6
>0.6–0.7
>0.7–0.8
>0.8–0.9
>0.9–1.0
>1.0–34.0 and N.A.
Gm07 8#1 0.52
Gm13 8#2 0.31
Gm18 8#2 0.41
Gm09 8#2 0.42
Gm09 8#9 0
Gm14 8#3 0.1
Gm07 8#3 0.52
Gm19 8#2 0.41
Gm10 8#7 1.81
Gm18 8#4 0.31
Gm16 8#1 0.41
Gm10 8#6 0.1
Gm14 8#7 0.62
Gm03 8#8 0.63
Gm16 8#2 0.73
Gm01 8#1 0.31
Gm20 8#2 0.1
Gm04 8#2 0.42
Gm18 8#1 0.21
Gm03 8#2 0.62
Gm03 8#5 0.1
Gm20 8#5 0.21
Gm05 8#2 0.73
Gm10 8#4 0.52
Gm14 8#1 0.15
Gm05 8#1 0.73
Gm11 8#1 0
Gm03 8#4 0.63
Gm16 8#4 0.54
Gm17 8#2 0.1
Gm16 8#5 1.05
Gm18 8#7 0.21
Gm02 8#4 0
Gm06 8#1 1.15
Gm18 8#5 0.21
Gm12 8#1 0.31
Gm14 8#4 0.41
Gm19 8#3 0.73
Gm02 8#3 N.A.
Gm19 8#4 0.41
Gm09 8#3 0.94
Gm15 8#4 0.21
Gm14 8#6 0.31
Gm01 8#2 0.75
Gm05 8#3 0.21
Gm07 8#2 0.52
Gm13 8#1 0.1
Gm05 8#4 1.22
Gm18 8#6 0.21
Gm10 8#3 0
Gm06 8#3 0.94
Gm10 8#5 0.73
Gm07 8#5 0.76
Gm11 8#2 0.94
Gm15 8#2 0.73
Gm20 8#3 0.41
Gm12 8#4 0.31
Gm09 8#4 0.42
Gm10 8#2 0.1
Gm09 8#6 0.1
Gm04 8#4 0.1
Gm06 8#4 0.31
Gm02 8#2 0.21
Gm09 8#7 0.11
Gm01 8#3 0.73
Gm03 8#7 0.41
Gm08 8#1 0.42
Gm19 8#1 0.41
Gm09 8#10 0.63
Gm06 8#2 0.31
Gm20 8#6 0.53
Gm18 8#3 0.52
Gm04 8#1 0.62
Gm10 8#1 0.42
Gm08 8#2 0.52
Gm09 8#8 0.52
Gm20 8#4 0.21
Gm02 8#1 0.31
Gm10 8#8 0.63
Gm12 8#2 0.21
Gm05 8#5 0.86
Gm04 8#3 0.41
Gm14 8#2 0.52
Gm01 273#2 2.64
Gm03 273#1 33.73
Gm11 273#1 5.09

Figure 15 Phylogenetic relationship among different *GmCOPIA78* elements in the Wm82 genome. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 20.02747252 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. The analysis involved 86 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 9089 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

2.5 Characterization of *GmCOPIA78* insertion in pericentromeric regions and chromosome arms

The average insertion time of Gmr8 and Gmr273, excluding 46 copies of solo LTRs, was 1.29 and 13.82 mya, respectively. In Gmr273, Gm03 273#1, had the insertion time 33.73 mya which was outstanding from the other that made the biased average insertion time than it used to be. Next, we estimated the effect of the located regions of each element by calculating the insertion time in pericentromeric regions and chromosome arms. There were 2.99 and 0.89 mya in pericentromeric regions (n = 75) and chromosome arms (n = 12), respectively. Therefore, the elements in chromosome arms were inserted more recently than that in pericentromeric regions. Moreover, the range of the insertion time in pericentromeric regions (0 to 33.73 mya) is wider than in chromosome arms (0.1 to 0.63 and 1.81 mya) (Figure 16A). The distribution of the insertion time in total (including both pericentromeric regions and chromosome arms) was most bunched up at the range 0.21 to 0.52 mya (Figure 16B).

To discover *GmCOPIA78* elimination from the genome, solo LTRs were analyzed because it was thought to be a major process for removal of LTR-RT DNA in plants via unequal homologous recombination (Devos et al., 2002). We used 87 copies of *GmCOPIA78* to identify solo LTRs. We identified 12 solo LTRs in Gmr8 family and 34 solo LTRs in Gmr273 family. In Gmr8, the number of solo LTRs was lower than intact elements about 1:7 ratio, while, the number of solo LTRs and intact elements were almost the same in Gmr273 family. Consequently, the ratio of solo LTRs to intact elements were calculated as 0.14 and 11.33 in Gmr8 and Gmr273, respectively (Table 1). In order to clarify the effect of chromosomal structure to the ratio of S/I, we investigated the structures of LTR-RTs separately in pericentromeric regions and chromosome arms. The S/I ratio in pericentromeric region was 0.41, whereas, it was 1.25 in chromosome arms. This result was also correlated to the previous

study in soybean, which reported lower S/I ratio in pericentromeric regions (1.26) than in chromosome arms (1.62) (Du et al., 2011).

Physical distance was also measured separately in pericentromeric regions and chromosome arms by measuring the distance from pericentromeric edge to left or right side depending on the location of the element based on centromere. In pericentromeric region, all group of the elements showed a slightly different average data and distribution pattern. When the time of insertion passed, the inserted copies gradually moved far from centromere to near pericentromeric edge. The average distance from pericentromeric edge of the period of 0-0.5, >0.5-1.0, and >1.0-34.0 & N.A. mya were about 7, 5, and 4 Mb, respectively. Solo LTRs that inserted into pericentromeric regions showed 5 Mb distance far from pericentromeric edge (Figure 17A). On the other hand, there were varieties of distribution patterns in chromosome arms. The elements inserted in the range >0.5-1.0 mya were distributed wider than the other, which ranging from about 1 to 7 Mb. In the range of >0.5-1.0 mya, the distribution was nearly similar to the elements inserted in the same period in pericentromeric regions. In addition, the distribution pattern of solo LTRs in chromosome arms also showed almost similar trend to those in pericentromeric regions. However, the elements inserted in the range 0-0.5 and >1.0-34.0 & N.A. mya contained 1 and 3 copies, respectively. Therefore, the distribution pattern might not be correct (Figure 17B). The real distribution pattern depending on each period was constructed by merging both distribution pattern in pericentromeric regions and chromosome arms determining minus and plus side to be the elements that were inserted into pericentromeric regions and chromosome arms, respectively. The average of all inserted period including solo LTRs were demonstrated that most of them were located in pericentromeric regions between 0 to 10 Mb from the pericentromeric edge (Figure 17C).

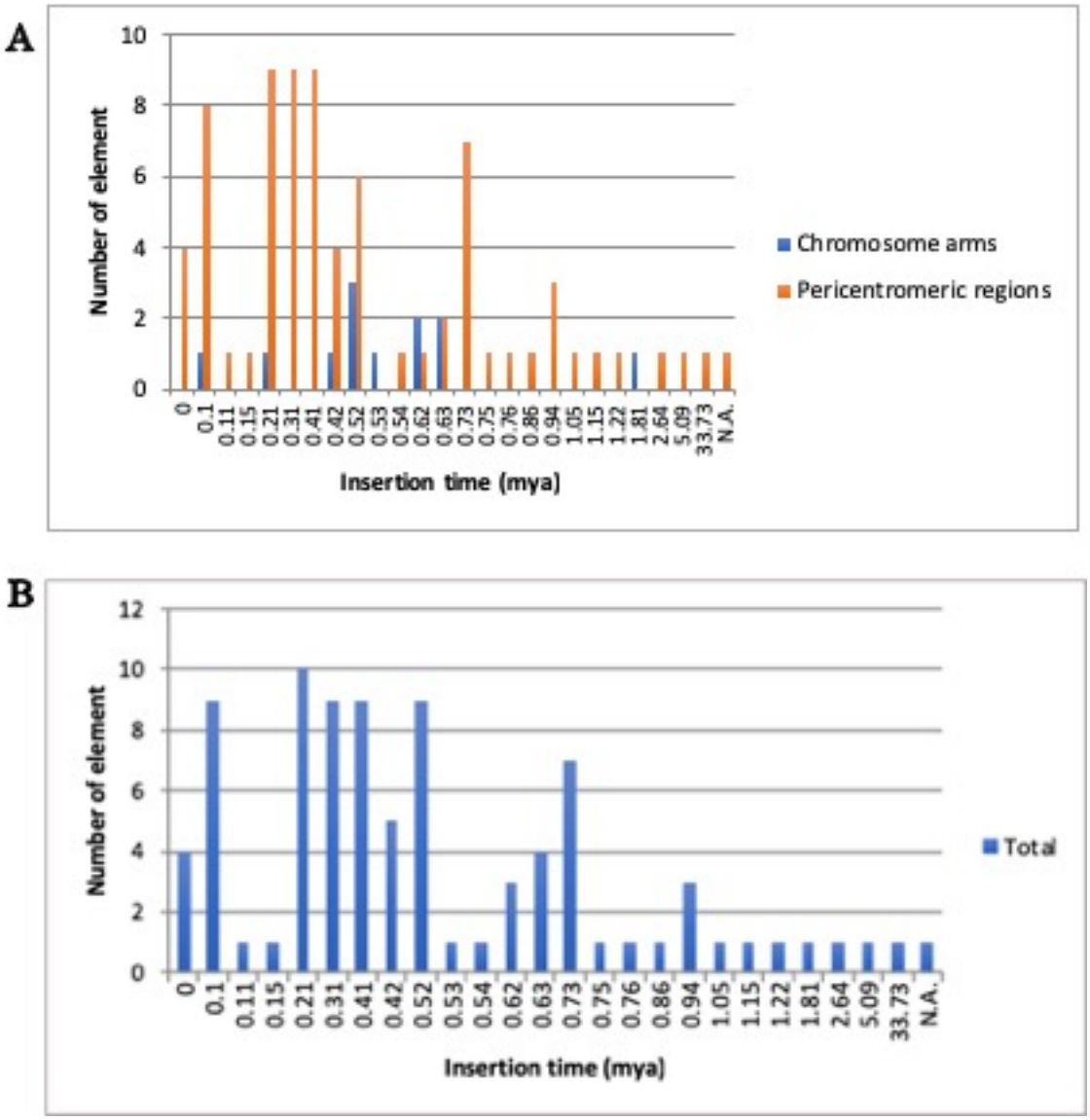

Figure 16 Number of *GmCOPIA78* elements compiled on the basis of insertion time.

A Insertions into chromosome arms (blue bar) and pericentromeric regions (orange bar). **B** Total insertion which included number of elements both in chromosome arms and pericentromeric regions.

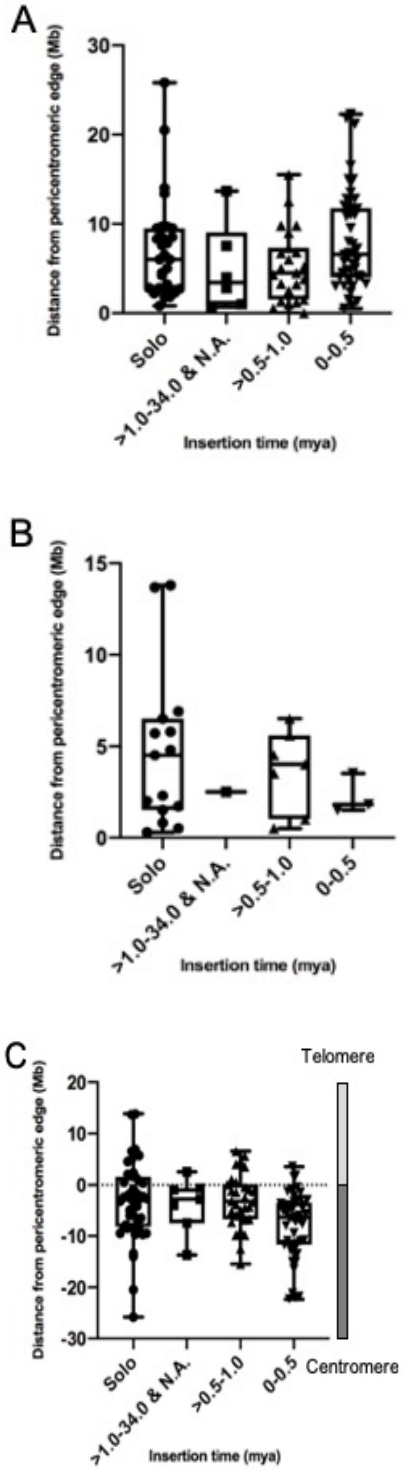

Figure 17 Distribution pattern of each *GmCOPIA78* element varying by insertion time.

A, B, and **C** Location of *GmCOPIA78* in pericentromeric regions (n=106), chromosome arms (n=27), and merging of both of them (n=133), respectively. N.A. (not analyzed). Data included with >1.0-34.0 insertion time group and solo LTR were also shown. For **C**, distance from the boundary between chromosome arms and pericentromeric regions within each region was

assigned by positive and negative value for elements located in chromosome arms and pericentromeric regions, respectively. The line and square within the box showed the median and mean value of all data that the bottom and top edges of the box indicated the 25th and the 75th percentiles of the data, respectively. The bottom and top bar indicated the minimum and maximum of the data, respectively.

2.6 Heat activation of *GmCOPIA78* in world varieties of soybean.

The previous studies showed that *ONSEN* (*COPIA78*) was conserved among Brassica species (Ito et al., 2013). In addition, the heat activation was conserved in adzuki bean which was in the same family (Masuta et al., 2017). To analyze the heat activation of *ONSEN*-like elements in soybean, I detected the extrachromosomal DNA of *GmCOPIA78* in the world varieties of soybean subjected to heat stress (Table 2). The extrachromosomal DNA was detected in 49.37% of analyzed varieties (Figure 18). Depend on the geographic regions, East Asia region contained the highest average percentage of detected ecDNA (55.45%). However, if the countries that had only one seed genotype was neglected, the highest and lowest percentage of ecDNA was detected in seed genotypes originated from India (77.8%) (n=9) and Pakistan (0%) (n=2), respectively (Table 3). It was remarkable that ecDNA of *GmCOPIA78* was not detected in *WILLIAM82* which is a source of analyzed data in Soybase.

ID	Name	Type	Origin
GmWMC125	BHATMAS	Wild	Nepal
GmWMC129*2	AOKI MAME	Wild	China
GmWMC132	L 2A	Wild	Philippines
GmWMC136	LOCAL VAR(SEPUTIH RAMAN)	Wild	Indonesia
GmWMC138	COL/PAK/1989/IBPGR/2326(1)	Wild	Pakistan
GmWMC141	PETEK	Wild	Indonesia
GmWMC142	JAVA 5	Wild	Indonesia
GmWMC143	M 44	Wild	India
GmWMC144	M 918	Wild	India
GmWMC146	HM 39	Wild	India
GmWMC147	COL/THAI/1986/THAI-78	Wild	Thailand
GmWMC148	M 42	Wild	India
GmWMC150	U 1042-1	Wild	Nepal
GmWMC151	JAVA 7	Wild	Indonesia
GmWMC152	U 1290-1	Wild	Nepal
GmWMC154	MANSHUU MASSHOKUTOU	Wild	China
GmWMC156	U 8006-3	Wild	Nepal
GmWMC159	COL/PAK/1989/IBPGR/2323(2)	Wild	Pakistan
GmWMC160*2	N 2392	Wild	Nepal
GmWMC162	COL/THAI/1986/THAI-80	Wild	Thailand
GmWMC163*2	N 2491	Wild	Nepal
GmWMC165	KARASUMAME(SHINCHIKU)	Wild	Taiwan
GmWMC166	MERAPI	Cultivated	Indonesia
GmWMC168	L 317	Wild	India
GmWMC169	HAKUCHIKOU	Wild	China
GmWMC170	M 652	Wild	India
GmWMC171	U-1741-2-2 NO.3	Wild	Nepal
GmWMC173	KARASUMAME(NAIHOU)	Wild	Taiwan
GmWMC175	BISHUU DAIZU	Wild	China
GmWMC176	SANDEK SIENG	Wild	Cambodia
GmWMC181	CHIENGMAI PALMETTO	Wild	Thailand
GmWMC182	LOCAL VAR(TEGINENENG)	Wild	Indonesia
GmWMC183	KARASUMAME(HEITOU)	Wild	Taiwan
GmWMC186	RINGGIT	Cultivated	Indonesia
GmWMC187	KADI BHATTO	Wild	Nepal
GmWMC188	E C 112828	Wild	India
GmWMC190	SAN SAI	Wild	Thailand
GmWMC191	MISS 33 DIXI	Wild	Philippines
GmWMC192	U 1155-4	Wild	Nepal

Table 2 List of world varieties of soybean analyzed to find *GmCOPIA78* ecDNA by Southern blotting. Genotypes are arranged based on GmWMC ID number. GmWMC: world soybean mini-core collections.

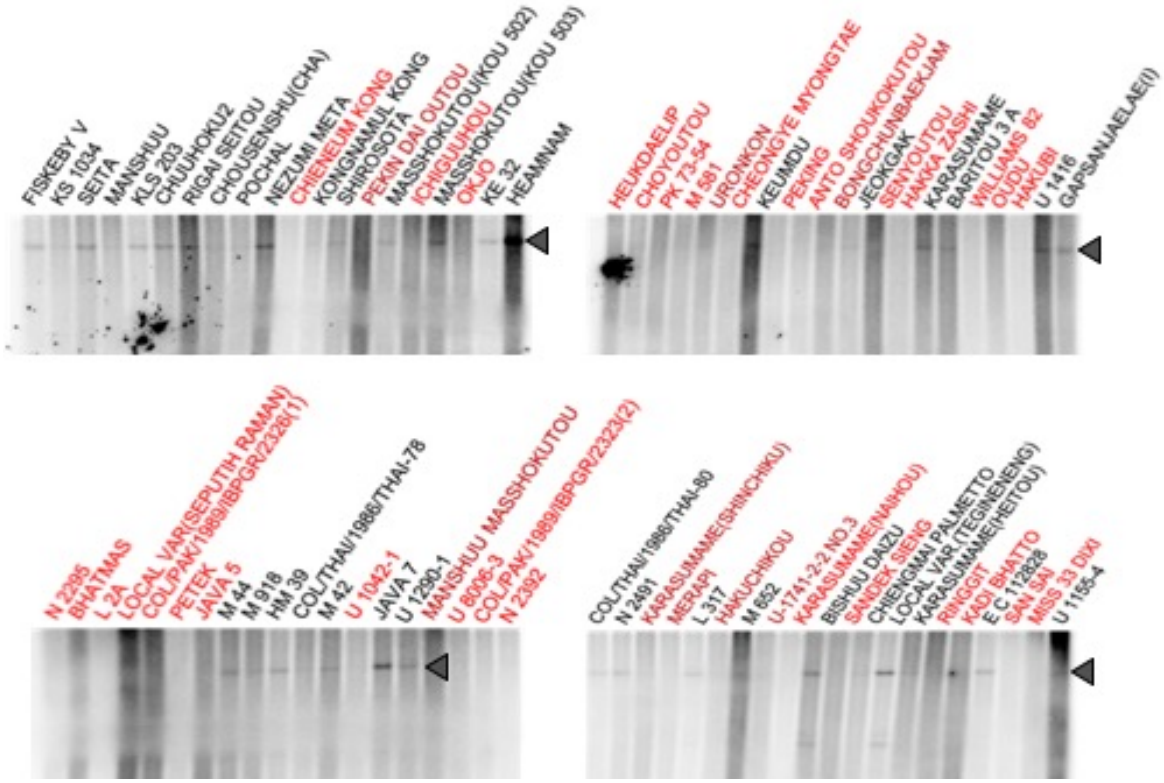

Figure 18 Southern blotting in world varieties of soybean after treating with heat stress. Red letters indicated no exDNA was detected. Arrows indicated extrachromosomal DNA of *GmCOPIA78*.

Region	Country	Seed genotype	Extrachromosomal DNA was detected		% of detected ecDNA
			Yes	No	
East Asia	China	17	6	11	35.3
	South Korea	13	8	5	61.5
	North Korea	4	3	1	75
	Taiwan	4	2	2	50
Southeast Asia	Philippines	3	1	2	33.3
	Indonesia	8	3	5	37.5
	Malaysia	1	1	0	100
	Thailand	4	3	1	75
	Cambodia	1	0	1	0
South Asia	Pakistan	2	0	2	0
	Nepal	11	4	7	36.4
	India	9	7	2	77.8
Northern America	USA	1	0	1	0
Northern Europe	Sweden	1	1	0	100

Table 3 Percentage of *GmCOPIA78* extrachromosomal DNA (ecDNA) in world varieties of soybean. It was arranged based on the 5 geographic regions of the world including East Asia, Southeast Asia, South Asia, Northern America, and Northern Europe.

2.7 The effect of Zebularine to the copy number of *GmCOPIA78*

A chemical inhibitor of DNA methyltransferases has been utilized to study the effect of DNA hypomethylation in plants (Pecinka & Liu, 2014) which may increase ecDNA. It has been determined that zebularine is more effective than other chemical demethylating agents (Baubec, Pecinka, Rozhon, & Mittelsten Scheid, 2009). To examine ecDNA of *GmCOPIA78* after treating with zebularine, qPCR was performed by using Kuro-goyo which is a commercial variety of soybean. The result showed that there was no significant difference among the all of heat-stressed samples (Figure 19).

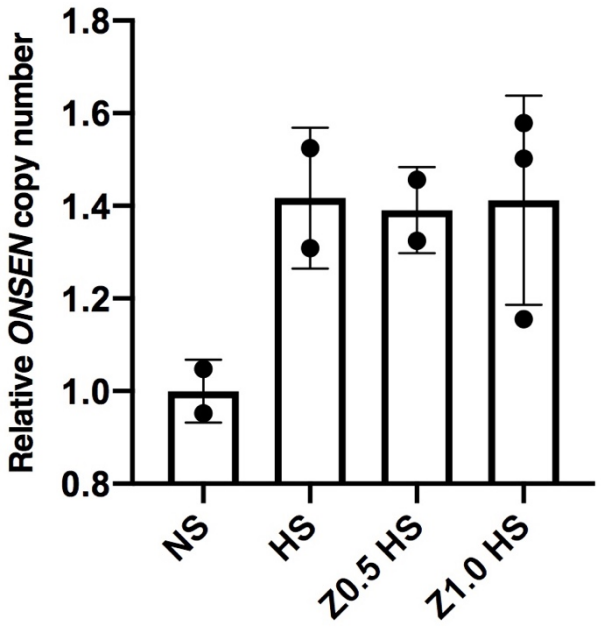

Figure 19 The relative copy number of *ONSEN* in the soybean (Kuro-goyo) seedlings subjected to heat stress (40°C for 24 hrs) comparing with the non-stress samples was analyzed by quantitative PCR (qPCR). Error bars indicate standard deviations (n=2-3). Ordinary one-way ANOVA for qPCR data were performed.

3 Discussion

3.1 The effect of zebularine for *ONSEN* activation in *Arabidopsis* and adzuki upon heat stress

When using zebularine to inhibit DNA methylation in the callus of *A. thaliana*, it was found that a concentration of 0.01 mM of zebularine could increase the copy number of *ONSEN*. However, 0.04 mM of zebularine did not induce more accumulation of *ONSEN* than that seen in the 0.01 mM treatment, meaning that *A. thaliana* contained a proper zebularine concentration that was required for *ONSEN* activation. A previous study detected somatic transposition of *ONSEN* in the calli of *Arabidopsis* and *Brassica* species after subjecting them to heat stress at 37 °C for 24 h (Ito et al., 2013). One of the benefits of callus-mediated transposition of *ONSEN* was that new alleles were generated, which thereby could help to expand the genetic variation of the tested plants. However, the presence of somaclonal variation was also detected, and this method requires relatively costly manipulations (Masuta et al., 2017). According to Masuta et al. (2018), the optimal heat stress condition for use in adzuki bean was 40 °C for 24 h, and extrachromosomal DNA was not detected in seedlings subjected to 37 °C for 24 h, which indicated that the appropriate temperatures for activating *ONSEN* vary among species. In addition, the appropriate concentration of zebularine to be use to activate *ONSEN* was also variable among species. Since, *A. thaliana* are smaller and more fragile than adzuki bean and soybean, the plant physical morphology might affect the endurable zebularine concentration. In case of *A. thaliana*, the *ONSEN* expression was highest at 0.01 mM zebularine. While in the proper concentration in adzuki bean was higher than *A. thaliana* around 50 folds, however, the right concentration of zebularine in soybean could not be identify.

The *ONSEN* expression level in adzuki bean does not always depend on the concentration of zebularine (Baubec et al., 2009). In the present study, the highest concentration did not result in the highest levels of *ONSEN* expression. The treatment with 0.5 mM of zebularine resulted in the highest *ONSEN* expression level, while the expression level

was decreased in the treatment with 1.0 mM of zebularine. Zebularine was previously found to be more stable in aqueous solutions and less cytotoxic than 5-azacytidine (Cheng et al., 2004). Zebularine has also been reported to inhibit DNA methyltransferases (e.g., DRM2), leading to decreased *ONSEN* DNA methylation, and subsequently activating the transcription of this element, resulting in the increased Pol II-dependent production of the sense and antisense transcripts (Thieme and Bucher, 2018).

According to Masuta et al., 2018, *ONSEN* was expressed in two varieties of adzuki bean, Tanba-Dainagon and Shumari. The different *ONSEN* expression pattern between Tanba-Dainagon and Shumari occurred, Shumari may be more sensitive to zebularine than Tanba-Dainagon because Shumari showed almost 6 folds higher than Tanba-dainagon after treating with 0.1mM zebularine. It was found that Shumari was one of the cultivars that insensitive to a long-day (16 h) conditions, in other words, it flowered under long day condition while other cultivars flowered normally delayed (Yamamoto et al., 2016). In addition, the 5-azacytidine which cause DNA demethylation could induce advanced flowering (Burn et al., 1993). It was thought to be involved in the regulatory mechanism of vernalization (Finnegan, 1998). In this experiment, the culture condition was also set to be a long day condition which was similar to a real culture condition in Hokkaido. Gathering both long day condition and zebularine treatment together lead Shumari to highly expression *ONSEN* even in low concentration of zebularine. However, both cultivars had their own optimal condition of zebularine treatment, since the expression level decreased when exceed concentration was applied. However, extrachromosomal DNA of *ONSEN* was only detected in the 'Tanba-dainagon' variety. Consequently, based on previous studies we applied different concentrations of zebularine to adzuki bean plants of the Tanba-dainagon subjected to heat stress. However, the 'Tanba-dainagon' variety had its own optimal zebularine concentration level for *ONSEN* expression, and the expression level decreased when a concentration exceeding this optimum one was

applied. The full-length extrachromosomal DNA detected was produced by assembly into VLPs. If there were no frame-shift mutations or stop codons in the CDS regions, this extrachromosomal DNA was able to integrate into a new locus or recombine with pre-existing elements (Servant and Deininger, 2016). The strength of the signal detected during Southern blotting was also the highest in the 0.5 mM treatment, and lower in that with 1.0 mM of zebularine, which consistent with the results found for *ONSEN* expression levels. The application of 5-azacytidine to seedlings and calli of *Sorghum bicolor* (L.) Moench demonstrated that the most suitable concentration of 5-azacytidine (310 μM) to induce TE expression in seedlings of this species was toxic in the callus (Emani et al., 2002). The optimal 5-azacytidine concentration was instead found to be 20 μM for the callus in that study. Therefore, the most suitable zebularine concentration for callus of adzuki bean to induce *ONSEN* activation should be sought out in the near future. Indeed, the extent of gene activation could be stronger in the callus than in seedlings because cell division activities are higher in callus tissues, resulting in faster rates of DNA synthesis (Weber et al., 1990). Treatment of plants with concentrations of DNA methylation inhibitors that are too high may lead to cell death because they have cytotoxic effects, such as by causing DNA damage via the covalent trapping of the DNA methyltransferase rather than DNA demethylation itself (Yamamoto et al., 2005).

Although zebularine was previously reported to be an efficient DNA methylation inhibitor, its effectiveness could be increased further by inhibiting RNA polymerase II (Pol II) via α-amanitin (Lindell et al., 1970). This is because the latter process would stabilize the active state of the TE by reducing the production of both its sense and antisense transcripts, which would decrease the accumulation of primary siRNAs, decrease post-transcriptional gene silencing, and increase retrotransposon extrachromosomal DNA production (Cuerda-Gil and Slotkin, 2016). Moreover, combining zebularine and α-amanitin together leads to the

production of massive amounts of extrachromosomal DNA, and resulted in the occurrence of 75 novel *ONSEN* insertions per plant in a previous study (Thieme et al., 2017).

In conclusion, the expression levels and production of extrachromosomal DNA of the heat-activated *ONSEN* retrotransposon in *A. thaliana* and *V. angularis* were found to be increased by treatment with zebularine. However, this increase did not occur in a dose-dependent manner, because each cultivar has its own optimal concentration that led to the highest expression levels of this retrotransposon in it. Our results should be useful in molecular breeding studies, for example to find more precocious strains of plants or those that could more readily adapt to environmental change such as global warming (Ito et al., 2016). If we could produce the new cultivars that can tolerate to higher temperature, agricultural harvested products will not lose as many as in the present situation. Currently, a method of using inhibitors of retrotransposon silencing exists, called BUNGEE (breeding using jumping genes) (Thieme et al., 2017), which is also useful in the production of non-transgenic crops and can promote molecular breeding by the use of new insertions, such as the retrotransposons examined in the present study. Applying this method with zebularine treatment should be helpful for discovering the new environmental stress tolerant cultivars in the near future.

3.2 Chromosomal distribution and effect of zebularine to heat-activated retrotransposon *ONSEN* in *Glycine max*

Among 133 copies of *ONSEN* in soybean, almost four-fifth of them were located in pericentromeric region leading to a transcriptionally repressive chromosomal environment and a high rate of DNA methylation (Chodavarapu et al., 2010). The RNAi pathway is supposed to be involved in these events, therefore, the expression of the repetitive elements is low in developmental stages and specific tissues (May et al., 2005). For example, there was a report in *Arabidopsis thaliana* that LTR retrotransposons were abundant in pericentromeric regions, therefore, genes that were located in pericentromeric regions were mainly deficient in their functions because it was strongly suppressed (Lin et al., 1999). Another factor to cause transcriptional degradation is the number of solo LTRs which they were found about 34.59% of *ONSEN* copies. The solo LTRs were formed by unequal intra-element homologous recombination resulting in removing of LTR retrotransposon DNA in plants (Devos et al., 2002). I supposed that when high amount of solo LTRs are detected, RdDM machinery might occur in lower rate than normal condition leading to decreasing in insertion of LTR retrotransposons to the new genomic loci.

After detecting chromosomal location of *ONSEN* in soybean, Gmr8 and Gmr273 families were identified. However, the elements in Gmr8 family inserted into the genome recently more than in Gmr273 and the copy of insertion into soybean genome was much higher. According to the data from Table 1, both of these families showed higher percentage of elements located in pericentromeric region than in chromosome arms. Nevertheless, most of the elements in Gmr273 family were solo LTRs in pericentromeric region, whereas, Gmr8 contained most intact elements. In addition, the data was supported by S/I (solo LTRs / Intact elements) ratio of Gmr273 (11.33:1), which it was much higher than Gmr8 (0.14:1). It can be also proved by comparing the average insertion time between these two families, they were

clarified that Gmr273 was inserted into the genome quite earlier than Gmr8. Therefore, Gmr273 family might gradually lost during the plant genome evolution because of the rapid elimination of retrotransposon DNA by unequal recombination such as solo LTRs formation (Devos et al., 2002). However, comparing with data from Du et al., 2010 that they characterized 510 LTR retrotransposon families in soybean demonstrated that most of the data was nearly similar to our data for Gmr8 but absolutely different for Gmr273. The biggest difference including number of solo LTRs (7 in Du's and 34 in my data), S/I ratio (0.70 in Du's and 11.33 in my data), and average insertion time (3.89 in Du's and 13.82 in my data). In our samples, the solo LTRs might be used to be intact elements over evolutionary period. It could be obviously distinguished between S/I ratio before the divergence to be different subspecies. For example, the S/I ratio of rice before the divergence to *Oryza indica* and *Oryza japonica* was 0.5 but was more than 1.0 after the divergence (Ma et al., 2004). I supposed the main cause of the difference that our source of sequence came from *A. thaliana COPIA78* (*AtCOPIA78*) sequence and blast the sequence through SoyBase while Du et al., 2010 used whole genome sequencing data directly from soybean cultivar Williams 82. Therefore, there might be some different in evolution between *AtCOPIA78* and *GmCOPIA78*. However, the S/I ratio of Gmr8 in pericentromeric regions (0.10) and chromosome arms (0.42) correlated to the previous result in soybean (Du et al., 2010) and rice (Tian et al., 2009) that both showed significantly lower S/I ratio in pericentromeric regions than in chromosome arms. Therefore, they suggested that the genetic recombination mechanisms for suppression in pericentromeric regions might decrease the solo LTRs formation frequency by unequal recombination (Du et al., 2010). According to Zamudio et al., 2015, they demonstrated that the elimination rate of TEs by unequal recombination was lower in pericentromeric or heterochromatic repetitive regions because a recombination rate was lower than in chromosome arms. In addition to unequal homologous recombination like solo LTRs, illegitimate recombination such as 3′ or 5′ LTR

deleted that generates small deletions was reported to be more effective LTR retrotransposon DNA elimination mechanism than unequal homologous recombination in *Arabidopsis* (Devos et al., 2002). However, our sequences did not contain only one side of LTR sequence, therefore, no need to include illegitimate recombination.

Depending on the insertion time of intact elements, the number of copies which was mostly inserted into the genome were estimated from 0.5 mya to the present. This result also correlated to previous results of *SORE-1* which is one of the other *Ty1/copia* retrotransposon in soybean that it was occurred from 0.1 mya to the present (Nakashima et al., 2018). The characteristic of these elements was suggested to be mutagens, because these elements were mostly transposed at the present time over their entire history (Nakashima et al., 2018). The distribution of the intact elements inserted in pericentromeric regions demonstrated that there were gradually shifted to locate near pericentromeric edge when the time of insertion passed. Therefore, the most recently inserted elements, 0-0.5 mya, showed the nearest location to centromere. While there was an opposite direction of the elements inserted in chromosome arms that the older elements might slightly shift to telomere. However, I did not involve >1.0-34.0 & N.A. group in this result interpretation for the elements located in chromosome arms because there was only one element in this group. Comparing to Nakashima et al., 2018's results, similar trend occurred in the elements in pericentromeric regions but opposite in chromosome arms. Indeed, the number of intact elements in chromosome arms were too low to find the elements shifted direction, therefore, it was ambiguous to identify. Vicient and Casacuberta, 2017 claimed that the regions around the centromeres and telomeres usually contain a higher TE density. It correlated to my results in that most elements inserted into the genome were estimated from 0.5 mya to the present, moreover, their inserted locations were the nearest to the centromeres. When TEs contain high density which many elements are gathering together in pericentromeric regions, the centromeres might have more ability to

endure microtubule tension during mitosis and meiosis (Freeling et al., 2015). In addition, the TEs insertion into the centromeres leads to the rapid evolution of the centromere, which is necessary for the species evolution (Han et al., 2016). For solo LTRs distribution, there was no bias between pericentromeric regions and chromosome arms which showed the same results to *SORE-1* (Nakashima et al., 2018). It can be implied that the elimination of both *GmCOPIA78* and *SORE-1* occurred in an unbiased manner. However, it will be better to use fluorescence in situ hybridization (FISH) method to confirm chromosome spreading location again.

After cultivating soybeans that have been collected depending on the geographic regions and subjecting to heat stress, it can be implied that the original cultivation area affected to the extrachromosomal DNA copy number. Even the highest extrachromosomal DNA percentage is 100% in Northern Europe region, it has only one genotype from one country. Therefore, the evidence was not enough. However, East Asia region was the interesting area because the detected extrachromosomal DNA percentage was 55.45%. The difference of detected *GmCOPIA78* signal might be LTR sequence mutation because LTR elements in *ONSEN* contains a HSFA2 heat-responsive element (Cavrak et al., 2014) and it normally binds to a cis-regulatory heat-responsive element resulting in transcription and production of full-length extrachromosomal DNA under elevated temperature (Ito et al., 2011). Therefore, losing HSFA2 function contribute to heat stress response depletion.

When applied zebularine treatment to Kuro-goyo variety, there was no difference among heat stress samples with 0, 0.5, 1.0 mM of zebularine treatment. Comparing to *Arabidopsis* and adzuki bean results, each plant has their own optimal zebularine concentration depending on size of plant because *Arabidopsis* could endure for low concentration of zebularine, whereas, adzuki bean and soybean showed a high zebularine concentration tolerance. However, the high concentration of zebularine should be limited not to exceed than 1.0 mM because it might affect to plant growth and lead to dwarfism (Figure 10). Therefore, it

was difficult to analyze extrachromosomal DNA copy number by using too much small seedlings. Soybean might not be a suitable model to study the effect of zebularine in seedlings. However, it is important to vary different concentration for zebularine treatment with soybean callus.

4 Materials and Methods

4.1 Plant materials, growing condition, and heat stress treatment

A wild-type *Arabidopsis thaliana* of the ecotype Columbia (Col-0), two varieties of adzuki bean (*Vigna angularis*) including Tanba-Dainagon, a commercially produced adzuki which was obtained from Takii & Company Limited, Kyoto, Japan and Shumari, which was derived from the Hokkaido Research Organization (Obihiro, Japan), and 79 world varieties of soybean (*Glycine max*) (Table 2) were used in this study. Seeds were grown on regeneration medium (RM) plates, containing 4.4 g l^{-1} MS plus vitamin (Sigma-Aldrich) with 0.5 g l^{-1} MES, 30 g l^{-1} sucrose, 4 g l^{-1} gellan gum, and 0.1 mg l^{-1} BA. One-week old seedlings were subjected to a temperature shift from 4°C for 24 hours to 37°C for 24 hours (in *A. thaliana*) or 40°C for 24 hours (in *V. angularis* and *G. max*). After the heat treatment, DNA and RNA were extracted from the seedling. As a control, seedlings were subjected to a temperature shift from 4°C for 24 hours to 21°C for 24 hours.

4.2 Callus induction

A. thaliana seedlings were plated on callus induction medium (CIM), containing 3.1 g l^{-1} Gamborg's B5 (Sigma-Aldrich) with 0.5 g l^{-1} MES, 20 g l^{-1} glucose, 2.5 g l^{-1} gellan gum, 0.05 mg l^{-1} kinetin and 0.5 mg l^{-1} 2,4-D, for 2 weeks. The induced callus was subjected to a temperature shift from 21 to 37°C for 24 hours. The heat-stressed callus was transferred to shoot induction medium (SIM), containing 3.1 g l^{-1} Gamborg's B5 with 0.5 g l^{-1} MES, 20 g l^{-1} glucose, 2.5 g l^{-1} gellan gum, 5 mg l^{-1} isopentenyl adenine and 0.1 mg l^{-1} IAA, after incubating for 2 weeks. The induced shoots were then transferred to root induction medium (RIM), containing 4.3 g l^{-1} MS with 0.5 g l^{-1} MES, 10 g l^{-1} sucrose, 2.5 g l^{-1} gellan gum, 2 mg l^{-1} 1-napthaleneacetic acid (NAA) and 1 ml l^{-1} 1,000 x B5, and the regenerated plants were transplanted to loose soil.

V. angularis seedlings were cut for 3-5 mm and plated on callus induction medium (CIM), containing 2.2 g l^{-1} MS plus vitamin (Sigma-Aldrich) with 0.5 g l^{-1} MES, 30 g l^{-1} sucrose, 4 g l^{-1} gellan gum, and 2 mg l^{-1} 2,4-D, with continuous light at 27°C for 2 weeks.

4.3 Chemical preparation

Zebularine (Tokyo Chemical Industry, Japan) was dissolved in water before being added to the MS medium at final concentration of 0.01 and 0.04 mM (in *A. thaliana*) or 0.1, 0.3, 0.5, and 1.0 mM (in *V. angularis* and *G. max*).

4.4 RNA extraction

Total RNA was extracted from the 7-day old seedlings using TRI Reagent (Sigma Aldrich, St. Louis, MO, USA), according to the manufacturer's instructions. Approximately 3-5 µg of the total RNA was treated with RQ1 RNase-free DNase (Promega, Madison, WI, USA). Reverse transcription was carried out using the ReverTraAce qPCR RT Kit (Toyobo, Osaka, Japan) with a random primer.

4.5 Southern blot analysis

Genomic DNA was isolated using the Nucleon PhytoPure DNA extraction kit (GE Healthcare Life Science, Chicago, IL, USA). Southern blotting was performed as described previously (Miura et al., 2004). Using a radiolabeled *VaONS*-specific probe generated using the Megaprime DNA Labeling System (GE Healthcare Life Science), in high sodium dodecyl sulfate hybridization buffer (Church & Gilbert, 1984), the DNA was digested with *Eco*RV and the hybridization signals were detected by P32 radio isotope.

4.6 Quantitative reverse transcription PCR (qRT-PCR) analysis

ONSEN gene-specific primers were designed based on candidate gene CDS sequence information (Table 4). 18S rRNA which is reported to be one the most stable housekeeping gene was designated to be the reference gene. The Luna Universal qPCR Master Mix (New England Biolabs, Inc., USA) and Applied Biosystems 7300 Real-Time PCR System (Applied Biosystems, Inc., CA, USA) were used to perform the qRT-PCR. The program used in this assay was as follows: 50°C for 2 min, 95 °C for 10 min, 95 °C for 15 s, 60°C for 1 min, 95 °C for 15 s, 60°C for 1 min, 95 °C for 15 s, 60°C for 15 s, 40 cycles. Three biological replicates were set for each sample.

Plant	Target/Purpose	Name	Primer
Arabidopsis	Southern probe	ONSEN-F	TAATGTTCCCTTCCAAGTCCC
		ONSEN-R	GCTTGTAATGACCCAAGAAGT
	qPCR	COPIA78_4129F	CCACAAGAGGAACCAACGAA
		COPIA_4300R	TTCGATCATGGAAGACCGG
		18Sr_FW	CGTCCCTGCCCTTTGTACAC
		18Sr-RV	CGAACACTTCACCGGATCATT
Adzuki bean	Southern probe	VaONS-F	ATGATGTCTGGGAGGTGGTC
		VaONS-R	CCCCCTCTTAAACTTGATTTTG
	qRT-PCR & qPCR	VaONS-F	GTCTGCCTTCTTAAATGGAGTTCTTGAA
		VaONS-R	AATCCGAACATTCCAAGCTCTCG
		qRT_VaUBC-F	AACAATTATGGGTCCTCC
		qRT_VaUBC-R	TGAAATACCTTCGTCCTG
		qPCR_Va18Sr_FW	CGTCCCTGCCCTTTGTACAC
		qPCR_Va18Sr-RV	GAACACTTCACCGGACCATT
Soybean	Southern probe	Gm02:35194637-F	GCCATGGACGAAGAAATCAA
		Gm02:35194637-F-R	ATTCCATCGGGGTGCCAA
	qPCR	Gmax_LTR_F	TGGAGGAGAGAACTTTGTATTTTTGC
		Gmax_LTR_R	TTCTTGTAGTGTTTGGAGTGTCT
		DmeACT-F	GACCTTCAACACCCCTGCT
		DmeACT-R	GTGGGAGTGCATAACCCTC

Table 4 Primer sequences used in these projects.

4.7 *ONSEN* chromosomal location, Solo LTRs identification, and phylogenetic relationship analysis

Eighty-seven LTR elements of *ONSEN* in the Williams 82 genome of soybean were analyzed by using Soybase (http://soybase.org) which is the database of soybean (Grant et al., 2010). First of all, we performed BLAST TE Database analysis by choosing the Transposable Elements from Genome Browser Tab and also selected a database to be a Williams 82 Transposable Elements. The blast highest score elements were selected to search about the details of those transposable elements. Finally, the start and end position were copied to see the region on the Genomic Map. Chromosomal location of each element was identified. For solo LTRs identification, the LTR *ONSEN* intact elements in the Williams 82 genome of soybean were used to identify the solo LTRs by using BLAT (BLAST-like Alignment Tool) of Soybase. For the phylogenetic relationship analysis, the MEGA7 software (Kumar et al., 2016) was used. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). Bootstrap values were calculated with 500 replicates.

References

Baubec, T., Pecinka, A., Rozhon, W. and Mittelsten, S. O. (2009). Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. Plant J, 57, 542-554.

Baubec, T., Finke, A., Mittelsten, S. O. and Pecinka, A. (2014). Meristem-specific expression of epigenetic regulators safeguards transposon silencing in *Arabidopsis*. EMBO Rep, Epub ahead of print.

Baucom, R. S., Estill, J. C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J. M., Westerman, R.P., SanMiguel, P.J. and Bennetzen, J. L. (2009). Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS genetics*, *5*(11), e1000732.

Beisler, J. A. (1978). Isolation, characterization, and properties of a labile hydrolysis product of the antitumor nucleoside, 5-azacytidine. J Med Chem, 21, 204-208.

Boothman, D. A., Briggle, T. V. and Greer, S. (1989). Exploitation of elevated pyrimidine deaminating enzymes for selective chemotherapy. Phamacol Ther, 42, 65-88.

Casacuberta, E. and Gonzalez, J. (2013). The impact of transposable elements in environmental adaptation. Mol. Ecol., 22, 1503-1517.

Cavrak, V. V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M. and Scheid, O. M. (2014). How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genet., 10, e1004115.

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

Champion, C., Guianvarc'h, D., Sénamaud-Beaufort, C., Jurkowska, R. Z., Jeltsch, A., Ponger, L., Arimondo, P.B. and Guieysse-Peugeot, A.-L. (2010). Mechanistic insights on the inhibition of c5 DNA methyltransferases by zebularine. *PloS one, 5*(8), e12388.

Chang, S. and Pikaard, C. S. (2005). Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation. J Biol Chem, 280, 796-804.

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P. Y., Stroud, H., Yu, Y., Hetzel, J. A., Kuo, F., Kim, J., Cokus, S. J., and Casero, D. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature*, *466*(7304), 388.

Church, G. M., & Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences, 81*(7), 1991-1995.

Coen, E. S., Robbins, T. P. and Almeida, J. (1989). Consequences and mechanisms of transposition in *Antirrhinum majus*. *Mobile DNA* (Ed. by D. E. Berg). & M. M. Howe, pp. 413–436. American Society for Microbiology, Washington DC.

Devos, K. M., Brown, J. K. and Bennetzen, J. L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome research*, *12*(7), 1075-1079.

Driscoll, J. S., Marquez, V. E., Plowman, J., Liu, P. S., Kelley, J. A. and Barchi, J. J. (1991). Antitumor properties of 2(1H)-pyrimidinone riboside (zebularine) and its fluorinated analogues. J Med Chem, 34, 3280-3284.

Du, J., Tian, Z., Hans, C. S., Laten, H. M., Cannon, S. B., Jackson, S. A., Shoemaker, R. C. and Ma, J. (2010). Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *The Plant Journal*, *63*(4), 584-598.

Duan, K., Ding, X., Zhang, Q., Zhu, H., Pan, A. and Huang, J. (2008). AtCopeg1, the unique gene originated from AtCopia95 retrotransposon family, is sensitive to external hormones and abiotic stresses. *Plant cell reports*, *27*(6), 1065-1073.

Feng, S., Jacobsen, S. E. and Reik, W. (2010). Epigenetic reprogramming in plant and animal

development. Science, 330, 622-627.

Fras, A., Juchimiuk, J., Siwinska, D. and Maluszynska, J. (2007). Cytological events in explants of *Arabidopsis thaliana* during early callogenesis. Plant Cell Rep., 26, 1933-1939.

Freeling, M., Xu, J., Woodhouse, M., and Lisch, D. (2015). A solution to the c-value paradox and the function of junk DNA: the genome balance hypothesis. Molecular Plant, 8, 899–910.

Gao, D., Abernathy, B., Rohksar, D., Schmutz, J., and Jackson, S. A. (2014). Annotation and sequence diversity of transposable elements in common bean (*Phaseolus vulgaris*). *Frontiers in plant science*, *5*, 339.

Goodier, J. L. and Kazazian Jr, H. H. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, *135*(1), 23-35.

Granbastien, M. A. (1998). Activation of plant retrotransposons under stress conditions. Trends Plant Sci., 3, 181-187.

Grant, D., Nelson, R. T., Cannon, S. B. and Shoemaker, R. C. (2010). Soybase, the USDA-ARS soybean genetics and genomics database. *Nucl. Acids Res*, 38(suppl 1), D843-D846.

Griffin, P. T., Niederhuth, C. E., and Schmitz, R. J. (2016). A comparative analysis of 5-azacytidine-and zebularine-induced DNA demethylation. *G3: Genes, Genomes, Genetics*, *6*(9), 2773-2780.

Han, J., Masonbrink, R. E., Shan, W., Song, F., Zhang, J., Yu, W., Wang, K., Wu, Y., Tang, H., Wendel, J.F. and Wang, K. (2016). Rapid proliferation and nucleolar organizer targeting centromeric retrotransposons in cotton. *The Plant Journal*, *88*(6), 992-1005.

Hurd, P. J., Whitmarsh, A. J., Baldwin, G. S., Kelly, S. M., Waltho, J. P. and Price, N. C. (1999). Mechanism-based inhibition of C5-cytosine DNA methyltransferases by 2-H

pyrimidinone. J Mol Biol, 286, 389-401.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature, 472*(7341), 115.

Ji, L., Neumann, D. A. and Schmitz, R. J. (2015). Crop epigenomics: identifying, unlocking, and harnessing cryptic variation in crop genomes. *Molecular plant, 8*(6), 860-870.

Jones, P. A. and Taylor, S. M. (1980). Cellular differentiation, cyditine analogs and DNA methylation. Cell, 20, 85-93.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16, 111-120.

Kirik, A., Pecinka, A., Wendeler, E. and Reiss, B. (2006). The chromatin assembly factor subunit FASCIATA1 is involved in homologous recombination in plants. Plant Cell, 18, 2431-2442.

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*, 33, 1870-1874.

Lall, I. and Upadhyaya, K. (2002). Panzee, a copia-like retrotransposon from the grain legume, pigeonpea (*Cajanus cajan* L.). *Molecular genetics and genomics, 267*(3), 271-280.

Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D., Fujii, C.Y., Mason, T., Bowman, C. L., Barnstead, M. and Feldblyum, T. V. (1999). Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. *Nature*, *402*(6763), 761.

Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K. and Abe, J. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the phytochome A gene. Genetics, 180, 995-1007.

LoSchiavo, F., Pitto, L., Giuliano, G., Torti, G. (1989). DNA methylation of embryogenic

carrot cell cultures and its variations as caused by mutation, differentiation, hormones and hypomethylating drugs. Theor. Appl. Genet., 77, 325-331.

Ma, J. and Bennetzen J. L. (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci U S A*, 101, 12404-12410.

Ma, J., Devos, K. M., & Bennetzen, J. L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research*, *14*(5), 860-869.

Marocco, A., Lorenzoni, C., & Francheboud, Y. (2005). Chilling stress in maize. *Maydica*, *50*(3/4), 571.

Masuta, Y., Nozawa, K., Takagi, H., Yaegashi, H., Tanaka, K., Ito, T., Saito, H., Kobayashi, H., Matsunaga, W. and Masuda, S. (2017). Inducible transposition of a heat-activated retrotransposon in tissue culture. Plant Cell Physiol., 58, 375-384.

Masuta, Y., Kawabe, A., Nozawa, K., Naito, K., Kato, A. and Ito, H. (2018). Characterization of a heat-activated retrotransposon in *Vigna angularis*. *Breeding science, 68*(2), 168-176.

Matzke, M. A. and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics, 15*(6), 394-408.

May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L., & Martienssen, R. A. (2005). Differential regulation of strand-specific transcripts from Arabidopsis centromeric satellite repeats. *PLoS genetics*, *1*(6), e79.

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

Miura, A., Kato, M., Watanabe, K., Kawabe, A., Kotani, H., & Kakutani, T. (2004). Genomic localization of endogenous mobile CACTA family transposons in natural variants of Arabidopsis thaliana. *Molecular genetics and genomics, 270*(6), 524-532.

Nakashima, K., Abe, J., & Kanazawa, A. (2018). Chromosomal distribution of soybean retrotransposon SORE-1 suggests its recent preferential insertion into euchromatic regions. *Chromosome Research*, *26*(3), 199-210.

Neelakandan, A. K. and Wang, K. (2012). Recent progress in the understanding of tissue culture-induced genome level changes in plants and potential applications. Plant Cell Rep., 31(4), 597-620.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A. and Schmutz, J. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, *457*(7229), 551.

Pecinka, A., Schubet, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J. and Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. Chromosoma, 113, 258-269.

Pecinka, A., & Liu, C. H. (2014). Drugs for plant chromosome and chromatin research. *Cytogenetic and genome research*,*143*(1-3), 51-59.

Pereira, V. (2004). Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome biology*, *5*(10), R79.

Rabinowicz, P. D., Palmer, L. E., May, B. P., Hemann, M. T., Lowe, S. W. and McCombie, W. R. (2003). Genes and transposons are differentially methylated in plants, but not in mammals. Genome Res., 13, 2658-2664.

Saariaho, A. H. (2006). Characterization of the molecular components and function of BARE-1, Hin-Mu and Mu transposition machineries.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4, 406-425.

Sakai, H., Naito, K., Ogiso-Tanaka, E., Takahashi, Y., Iseki, K., Muto, C., Satou, K., Teruya,

K., Shiroma, A., Shimoji, M. and Hirano, T. (2015). The power of single molecule real-time sequencing technology in the de novo assembly of a eukaryotic genome. *Scientific reports, 5*, 16780.

SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. and Bennetzen, J. L. (1998). The paleontology of intergene retrotransposons of maize. *Nat Genet*, 20, 43-45.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J.J., Cheng, J. and Xu, D. (2010). Genome sequence of the palaeopolyploid soybean. *nature*, *463*(7278), 178.

Schulman, A. H. (2012). Hitching a ride: nonautonomous retrotransposons and parasitism as a lifestyle. In *Plant Transposable Elements* (pp. 71-88). Springer, Berlin, Heidelberg.

Schulman, A. H. (2013). Retrotransposon replication in plants, Curr. Opin. Virol., 3, 604–614.

Shiraishi, M., & Hayatsu, H. (2004). High-speed conversion of cytosine to uracil in bisulfite genomic sequencing analysis of DNA methylation. *DNA research, 11*(6), 409-415.

Shoemaker, R. C., Schlueter, J., & Doyle, J. J. (2006). Paleopolyploidy and gene duplication in soybean and other legumes. *Current opinion in plant biology*, *9*(2), 104-109.

Soppe, W. J. J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M. S., Jacobsen, S. E., Schubert, I. and Fransz, P. F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J., 21, 6549-6559.

Stround, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V. and Jacobsen, S. E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. Cell, 152, 352-364.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the

sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22, 4673-4680.

Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J. L., Jackson, S. A., Gaut, B.S. and Ma, J. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons?. *Genome research*, *19*(12), 2221-2230.

Tian, Z., Zhao, M., She, M., Du, J., Cannon, S. B., Liu, X., Xu, X., Qi, X., Li, M.W., Lam, H.M. and Ma, J. (2012). Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. *The plant cell*, *24*(11), 4422-4436.

Vicient, C. M., and Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Annals of Botany*, *120*(2), 195-207.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O. and Paux, E. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973.

Woodrow, P., Pontecorvo, G., Ciarmiello, L. F., Fuggi, A., and Carillo, P. (2011). Ttd1a promoter is involved in DNA–protein binding by salt and light stresses. *Molecular biology reports*, *38*(6), 3787-3794.

Xu, Y., and Du, J. (2014). Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (*Solanum lycopersicum*) plants. *The Plant Journal*, *80*(4), 582-591.

Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., and Bourc'his,

D. (2015). DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes & development*, *29*(12), 1256-1270.

Zhang, X., Yazaki, J., Sundaresan, S., Cokus, S. and Chan, S. W. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell, 126, 1189-1201.

Zhao, C., Takeshima, R., Zhu, J., Xu, M., Sato, M., Watanabe, S., Kanazawa, A., Liu, B., Kong, F., Yamada, T. and Abe, J. (2016). A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. *BMC plant biology*, *16*(1), 20.

Acknowledgements

I would like to express my gratitude to professor Atsushi Kato and associate professor Hidetaka Ito for technical supporting and manuscript proofreading and all laboratory members of Cell Function and Structure III for technical supporting and discussion of my study. I would like to thank you to professor Tomomichi Fujita and professor Junji Yamaguchi for being my doctoral thesis committee and suggesting many valuable comments for my thesis. I can't thank you enough to associate professor Akira Kanazawa for recommending about chromosomal location study in soybean, associate professor Akira Kawabe for providing whole genome sequencing data of *AtCOPIA78* LTR region, and Dr. Hitoshi Satou for giving the seeds of adzuki bean (Shumari variety). Finally, I would like to thank MEXT Scholarship: IGP Special Quota for granting me a great opportunity to study in Biosystems Science Course, Graduate School of Life Science, Hokkaido University.