Repeated evolution of heat responsiveness among *Brassicaceae* *COPIA* transposable elements

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Björn Pietzenuk

aus Duisburg, Deutschland

Köln 2015

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln in der Abteilung für Pflanzenzüchtung und Genetik (Prof. Dr. Maarten Koornneef) in der Arbeitsgruppe von Dr. Ales Pecinka angefertigt.

Berichterstatter: Prof. Dr. Maarten Koornneef

Prof. Dr. Achim Tresch

Tag der mündlichen Prüfung: 29.10.2015

„Live is not about
waiting for the storm to pass,
It´s about learning
how to dance in the rain.”
– George Gregory “Greg” Plitt Jr. (1977 – 2015)

Abstract

Eukaryotic genomes contain repetitive DNA sequences. This includes simple repeats and more complex transposable elements (TEs). Many TEs reach high copy numbers in the host genome, owing to their amplification abilities by specific mechanisms. There is growing evidence that TEs contribute to gene transcriptional regulation. However, excess of TE activity may lead to reduced genome stability. Therefore, TEs are suppressed by the transcriptional gene silencing machinery via specific chromatin modifications. In contrary, effectiveness of the epigenetic silencing mechanisms imposes risk for TE survival in the host genome. Therefore, TEs may have evolved specific strategies for bypassing epigenetic control and allowing the emergence of new TE copies. Recent studies suggested that the epigenetic silencing can be, at least transiently, attenuated by heat stress in *A. thaliana*. Heat stress induced strong transcriptional activation of *COPIA78* family LTR-retrotransposons named *ONSEN*, and even their transposition in mutants deficient in siRNA-biogenesis. *ONSEN* transcriptional activation was facilitated by the presence of heat responsive elements (HREs) within the long terminal repeats, which serve as a binding platform for the HEAT SHOCK FACTORs (HSFs).

This thesis focused on the evolution of *ONSEN* heat responsiveness in *Brassicaceae*. By using whole-transcriptome sequencing approach, multiple *Arabidopsis lyrata ONSENs* with conserved heat response were found and together with *ONSENs* from other *Brassicaceae* were used to reconstruct the evolution of *ONSEN* HREs. This indicated ancestral situation with two, in palindrome organized, HSF binding motifs. In the genera *Arabidopsis* and *Ballantinia*, a local duplication of this locus increased number of HSF binding motifs to four, forming a high-efficiency HRE. In addition, whole transcriptome analysis revealed novel heat-responsive TE families *COPIA20*, *COPIA37* and *HATE*. Notably, *HATE* represents so far unknown *COPIA* family which occurs in several *Brassicaceae* species but is absent in *A. thaliana*. Putative HREs were identified within the LTRs of *COPIA20*, *COPIA37* and *HATE* of *A. lyrata*, and could be preliminarily validated by transcriptional analysis upon heat induction in subsequent survey of *Brassicaeae* species. Subsequent phylogenetic analysis indicated a repeated evolution of heat responsiveness within *Brassicaceae COPIA* LTR-retrotransposons. This indicates that acquisition of heat

responsiveness may represent a successful strategy for survival of TEs within the host genome.

Zusammenfassung

Eukaryotische Genome beinhalten sich wiederholende DNA Sequenzen. Dies umfasst einfache Sequenzwiederholungen und die komplexeren Transposons. Aufgrund ihrer Fähigkeit sich durch spezifische Mechanismen selbst zu vervielfältigen, erreichen viele Transposons eine hohe Kopienzahl innerhalb eines Genoms. Es gibt immer mehr Hinweise darauf, das Transposons auch einen Einfluss auf die transkriptionelle Genregulierung haben können. Jedoch kann die Aktivität von Transposons zu einer verringerten Genomstabilität führen. Deshalb werden Transposons durch transkriptionelle Gen-Stilllegung (TGS) mittels spezifischer Chromatinmodifikationen gehemmt. Dahingegen führt die Effektivität der epigenetischen Stilllegung zu einer Gefährdung des Überlebens eines Transposons im Wirtsgenom. Deswegen haben Transposons möglicherweise spezifische Strategien entwickelt um die epigenetische Kontrolle zu umgehen, das Ihnen erlaubt neue Kopien zu erstellen. Kürzlich veröffentlichte Studien in *A. thaliana* deuteten an, dass die epigenetische Stilllegung, zumindest vorrübergehend, durch Hitzestress geschwächt werden kann. Hitzestress induziert eine starke transkriptionelle Aktivität der LTR-retrotransposons aus der *COPIA78* Familie namens *ONSEN* und führt sogar zu dessen Transposition in siRNA Biosynthese Mangelmutanten. Die transkriptionelle Aktivierung von *ONSEN* wurde ermöglicht durch die Anwesenheit von sogenannten „heat responsive elements“ (HRE) innerhalb der langen terminalen Sequenzwiederholungen (LTR), die als Bindeplattform dienen für Hitzeshock-Transkriptionsfaktoren (HSF).

Diese Doktorarbeit beschäftigt sich mit der Evolution der durch Hitze induzierten Aktivierung von *ONSEN* in *Brassicaceae.* Durch die Verwendung der „Gesamt-Transkriptom-Shotgun-Sequenzierung“-Methode wurden mehrere *ONSEN* mit konservierter Hitzeinduzierbarkeit in *Arabidopsis lyrata* gefunden und zusammen mit *ONSEN* von anderen Brassicaceae dazu verwendet die Evolution von *ONSEN* HRE zu rekonstruieren. Dies deutete auf eine ursprüngliche Situation mit zwei, zu einem Palindrom angeordneten, HSF-Bindemotiv hin. Eine lokale Duplikation dieses Sequenzabschnittes, in den Genera *Arabidopsis* und *Ballantinia*, erhöhte die Anzahl der HSF-Bindemotive, was zur Ausprägung eines HRE mit hoher Bindungseffizienz führte. Zusätzlich konnte durch die Analyse der Gesamt-Transkriptom-Shotgun-

Sequenzierung, die neuen hitzeinduzierbaren Transposonfamilien *COPIA20*, *COPIA37* und *HATE* gefunden werden. Insbesondere *HATE* repräsentiert dabei eine neue bisher unbekannte *COPIA*-Familie, welche in einigen Brassicaceae auftritt, aber in *A. thaliana* fehlt. Mögliche HRE wurden in den LTRs von *COPIA20*, *COPIA37* und *HATE* in *A. lyrata* entdeckt und konnten bei anschließender Prüfung in *Brassicaceae* Arten vorläufig durch eine Transkriptionsanalyse auf Hitzeinduzierbarkeit bestätigt werden. Eine anschließende phylogenetische Analyse deutete auf eine wiederholte Evolution der Hitzeinduzierbarkeit von *COPIA* LTR-Retrotransposons innerhalb der Brassicaceae hin. Dies wiederrum deutet darauf hin, dass der Erwerb von Hitzeinduzierbarkeit eine erfolgreiche Strategie zum Überleben von Transposons im Wirtsgenom ist.

Table of Contents

1 Introduction

1.1 What are transpoable elements?

Transposable elements (TEs) are typically multi- to high-copy genetic elements that can change their location within the genome by transposition. They can be autonomous or non-autonomous (Figure: 1.1). While autonomous elements contain all necessary components for active transposition, non-autonomous ones rely on the presence of proteins from autonomous TEs to transpose.

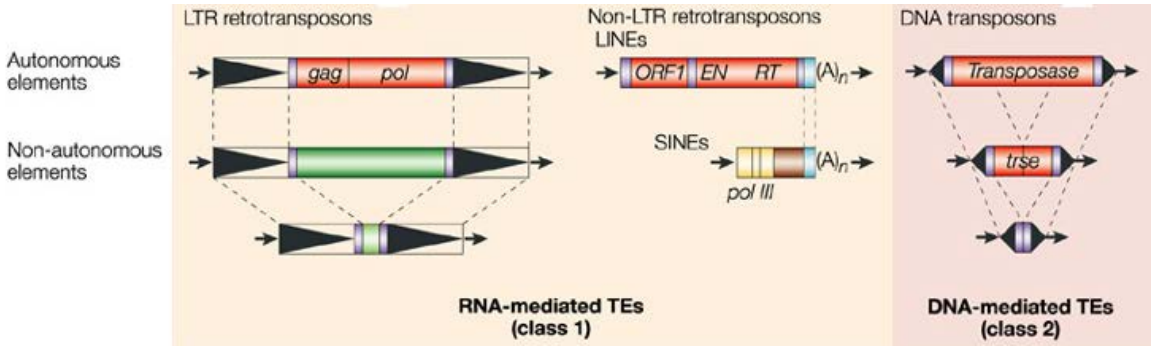

Figure 1-1: Structure of autonomous and non-autonomous transposable elements within Class I and Class II (modified after Feschotte, 2002 #85).

TEs were discovered by Barbara McClintock in the 1940s as “jumping genes” in maize (McClintock, 1950). She found that there are mobile DNA elements in the maize genome which can translocate within the genome and lead to phenotypic differences by jumping into genes encoding for a visible character e.g. crop pigmentation. Owing to their primarily self-amplification behavior TEs were considered as “junk DNA” (Ohno, 1972) or “genomic parasites”(Doolittle et al., 1980; Orgel et al., 1980). However, there is growing evidence that TEs can play an important role in regulation of gene transcription and are driving genome evolution (Tenaillon et al., 2010; Lisch, 2013b).

Transposable elements are divided into two classes, based on their transposition mechanism: “copy and paste” (Class I or Retrotransposons) or “cut and paste” (Class II or DNA-Transposons). This classification was proposed by Finnegan (1989) and later revised and extended by Wicker et al. (2007). A recent publication proposed a Nuclease/Recombinase based TE classification, because of the abundance of new TE findings notably within prokaryotes (Piégu et al., 2015). However, the focus of this

thesis lies on the evolution of stress responsiveness in *Ty1/COPIA*-retrotransposons, therefore, the well-established classification (Wicker et al. (2007) will be used.

1.1.1 Class I elements

Class I elements or retrotransposons are characterized by their transposition mechanism, which is known as “copy and paste”. They are amplified within the genome by an RNA intermediate, which will be converted into complementary DNA by a TE encoded reverse transcriptase (RT) (Wicker et al., 2007). These new DNA copies can be integrated into the genome by a TE encoded Integrase (INT). Due to this transposition mechanism, leading to direct amplification of the TEs within the genome, retrotransposons can constitute a large fraction in eukaryotic genomes (Wicker et al., 2007).

Retrotransposons are divided into 5 subclasses: LTR-retrotransposons, *DICTYOSTELIUM* INTERMEDIATE *REPEAT* (*DIR*)-sequence, *PENELOPE-LIKE ELEMENTS* (*PLE*), *LONG INTERSPERSED NUCLEOTIDE ELEMENTS* (*LINE*) and *SHORT INTERSPERSED NUKLEOTIDE ELEMENTS* (*SINE*) (Wicker et al., 2007). LTR-retrotransposons form the most abundant TE superfamily within large plant genomes (Mao et al., 2000; Feschotte et al., 2002; Lisch, 2009; Tenaillon et al., 2010; Fedoroff, 2012; Lisch, 2013a). They are characterized by a 3´and 5´flanking repetitive sequence called long terminal repeats (LTRs). LTR´s are the transcriptional regulators for the LTR-retrotransposons. While the 3´LTR works as a promotor, the 5´LTR takes part as a transcription terminator (Casacuberta et al., 2003). Adjacent to the 3´end of the 5´LTR resides the primer binding site (PBS) which initiates reverse transcription (Havecker et al., 2004), while a polypurine tract previous to the 3´LTR prevents mRNA from digestion by RNase H and initiates plus-strand synthesis (Rausch et al., 2004). The coding region of LTR-retrotransposons is divided into two domains, transcribed from a single open reading frame (ORF). These two domains are: GAG, which encodes for a Virus-like particle (VLP) and POL, which encodes for a reverse transcriptase (RT), integrase (INT), aspartic protease (AP) and family specific RNAse H (RH) (Figure: 1-2).

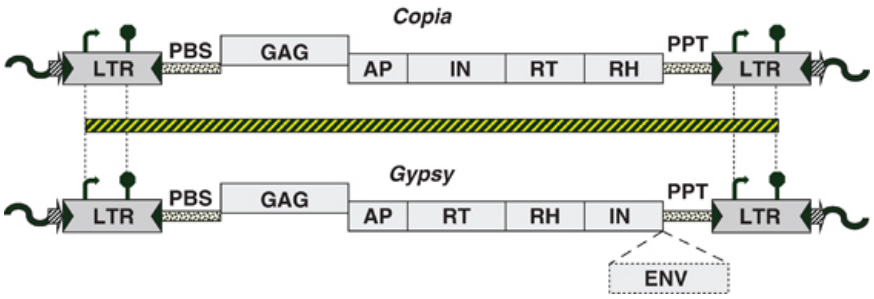

Figure 1-2: Organization of plant LTR-retrotransposons. LTR-retrotransposons in plants are divided into *Gypsy* and *COPIA* as shown. LTR-retrotransposons are bound by long terminal repeats (LTRs). Target side duplications are marked as small arrows at the LTR ends. Reverse transcription is primed at the PBS and PPT domains. The transcript is indicated by a hatched box between the *COPIA* and *Gypsy* diagrams. The proteins needed for a retrotransposon life-cycle are divided into two domains and is transcribed from a single open reading frame (ORF). The GAG-Domain encodes for a Virus-like particle, while the POL-domain encodes for a aspartic protease (AP), integrase (INT), reverse transcriptase (RT) and family specific RNAse H (RH). An additional ORF of an Envelope protein (ENV) found in some Gypsy elements is indicated. Waved lines at the ends mark adjacent genomic DNA (modified after (Kalendar et al., 2011).

These proteins are needed for transposition of LTR-retrotransposons. LTR-retrotransposons are divided into 5 superfamilies, which can be distinguished by their LTR sequences in length and composition as well as the alignment of the POL-domain. The two largest superfamilies within plant species are *Gypsy* and *COPIA*, which differ in the order of the integrase enzyme. While in *COPIA* TEs, the Integrase is upstream of the POL domain, it is downstream in the POL domain of *Gypsy* elements (Wicker et al., 2007). However, all LTR-retrotransposons transpose by the same mechanism. The transposition mechanism of LTR-retrotransposons was described by Hirochika (1993) and Böhmdorfer et al. (2005) on the basis of *Tto1* retrotransposon from tobacco. *Tto1* is an autonomous LTR-retrotransposon of the *COPIA* superfamily with a total length of 5.3 kb, flanked by LTRs with an average size of 574 bp. Transposition of *Tto1* starts with the transcription of the elements from the 5´LTR mediated by RNA polymerase II. This transcript serves as mRNA, encoding AP, VLP, RT, RH and INT as well as being template for reverse transcription. The translated Polyprotein assembles in the VLP encoded by the GAG-domain, where the AP cleaves it to release the INT, RT and RH. After reverse transcription of the mRNA by the RT, arisen cDNA is bound by the INT and inserted at a new genomic location (Figure: 1-3). Thus a new, identical copy of the original element is produced. Non-autonomous LTR-retrotransposons lack partially or totally

of their protein coding region, but mostly still contain LTR´s PBS and PPT which are minimal requirements to initiate reverse transcription (Havecker et al., 2004).

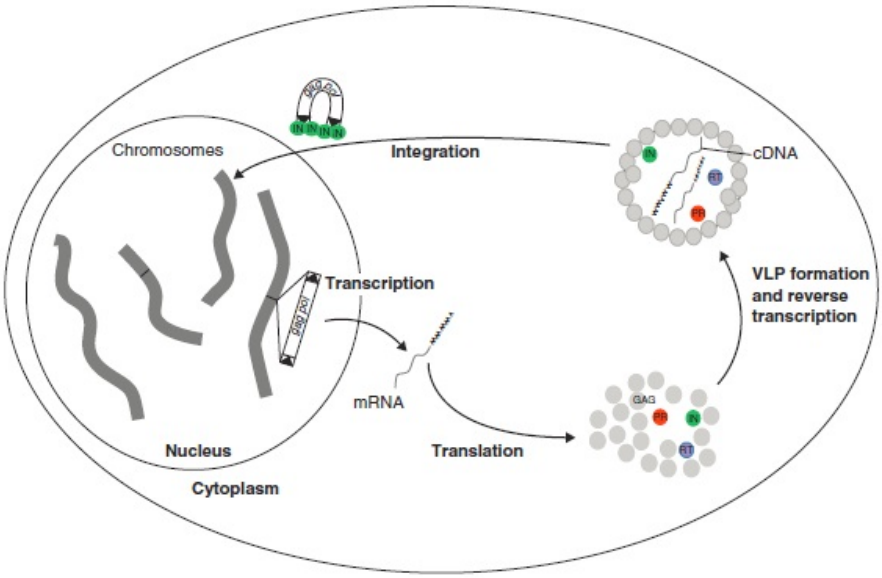

Figure 1-3: Life cycle of a LTR-retrotransposon. GAG = Capsid-protein; IN = Integrase; RT = reverse Transcriptase; PR = Protease (modified after Havecker, 2004 #105).

Beside *Gypsy* and *COPIA*, there are 3 more superfamilies within the subclass of LTR-retrotransposons: *BEL-Pao*, *Retroviruses* and endogenous retroviruses (*ERV*). They are so far all restricted to Metazoans (Wicker et al., 2007). *Retroviruses* and *ERV* are structurally and phylogenetically close to *Gypsy* LTR-retrotransposons, containing a GAG-POL-Domain and flanking LTRs. However, they encode additionally a viral envelope protein (ENV), which indicates a viral lifestyle (Frankel et al., 1998; Seelamgari, 2004). *BEL-Pao* elements are similar to *COPIA* and *Gypsy* elements, building an own clade based on RT phylogenies (Xiong et al., 1993; Cook et al., 2000; Wicker et al., 2007).

Concerning the other four subclasses within Class I transposons. *DIR*s contain a reverse transcriptase, but integrate in the genome via a tyrosine recombinase (YR) except of an integrase (Cappello et al., 1985; Goodwin et al., 2004a). *PLE*s encode a telomerase related reverse transcriptase and transposes by an endonuclease (Evgen'ev, 1997; Evgen'ev et al., 2005). *LINE*s do not contain LTRs, but encode at least a reverse transcriptase and a nuclease for transposition. *SINE*s are small non-autonomous elements that rely on *LINE* for transposition such as reverse transcriptase (Kajikawa et al., 2002; Dewannieux et al., 2003; Kramerov et al., 2005; Wicker et al., 2007).

1.1.2 Class II elements

Class II elements or DNA transposons are characterized by the “cut and paste” transposition mechanism. They consists of two major subclasses which are differentiated by the number of cut DNA strands while amplification. Subclass I includes on one hand of TIR (terminal inverted repeat) TEs which are defined by a transposase containing coding region flanked by terminal inverted repeats at the 3´and 5´end. Their transposition is mediated by a self-encoded transposase, which recognizes the TIRs and cuts at 3´and 5´ends (Figure 1-4 A). Thereby, the transposase produces overhangs (sticky ends) which will be gap repaired resulting into target site duplication (TSD). All nine families within subclass one are distinguished by the TIR sequence and their TSD (Wicker et al., 2007). In addition, a second order consists to subclass II called *Crypton* (Goodwin et al., 2003). They were currently exclusively found in fungi and lack of TIRs and a transposase domain. But instead, they encode a YR and also generate TSDs. Transposition of *Crypton* TEs is proposed as recombination between a circular molecule and a DNA target, requiring cleavage of both DNA strands (Goodwin et al., 2003; Goodwin et al., 2004b; Wicker et al., 2007). Thereby, they produce TSD, but they are lacking of TIRs.

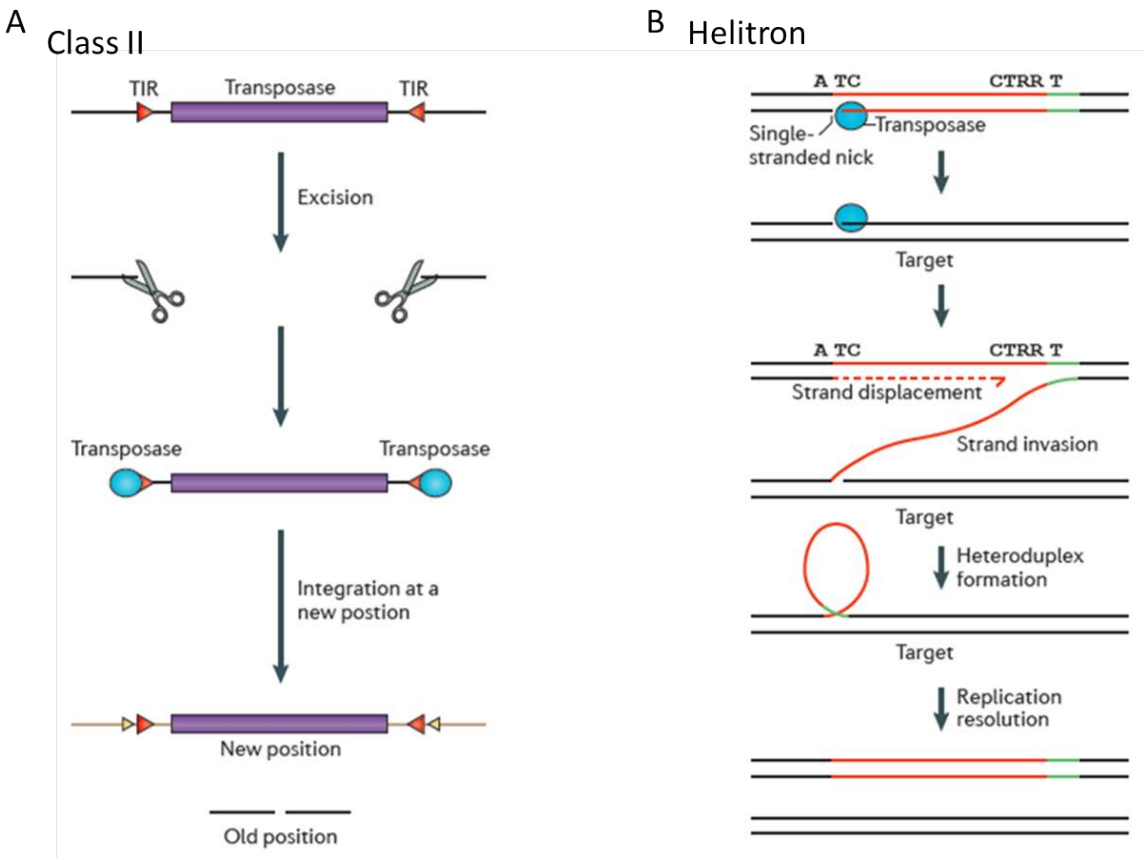

Figure 1-4: Transposition of (A) Class II and (B) Helitron DNA transposons (modified after (Lisch, 2013a)).

Subclass 2 consists of *Helitron and Maverick* TEs. These are DNA Transposons that transpose without double-strand cleavage. *Helitron* TEs transpose via a rolling-circle mechanism where only one strand will be cut from the host sequence and re-integrated within the genome (Figure 1-4 B; (Kapitonov et al., 2001)). *Helitron* TEs are best described in the maize genome, which contain a lot of non-autonomous derivatives (Kapitonov et al., 2001; Wicker et al., 2007). They are characterized by their TC or CTRR motifs at their ends (where R is A or G) and a short hairpin hairpin-structure before the 3´end. Autonomous *Helitron* encodes a Y2-type tyrosine recombinase (YR), close to that of the bacterial *IS91* rolling-circle transposons, with a helicase domain and replication initiator activity.

Maverick (a.k.a. *Polintons*) TEs are rarely present in eukaryotic genomes and were not found in plant genomes to date (Pritham et al., 2007; Wicker et al., 2007). They are relatively large (10-20 kb), feature TIRs at their ends and encode up to 11 proteins which differ in number and order. *Maverick* TEs are proposed to transpose by single strand excision, followed by extrachromosomal replication with subsequent

new integration, which is called replicative transposition without RNA intermediates. This assumption gets underlined by the fact, that *Mavericks* TEs encode a DNA polymerase B and an Integrase (related to the Class I type) but lack of a reverse Transcriptase (Kapitonov et al., 2006; Wicker et al., 2007).

1.2 Transposable elements in plants

The “selfish” and “parasitic” behavior of TEs was used to explain the C-value paradox (Gall, 1981). The C-value paradox is the extensive variation in nuclear DNA content in eukaryotic genomes, which does not correlate number of genes. Actually, the large variance in the DNA content of flowering plants is a perfect example for the C-value paradox (Figure 1-5 A; (Fedoroff, 2012)). Interestingly, the genome size differ in angiosperms by ~2100 fold, and correlate strongly with the TE content (Figure: 1-5 B; (Vitte et al., 2006; Gregory et al., 2007; Tenaillon et al., 2010). In contrast to this, the number of genes remains relatively the same (Gregory, 2005).

Small plant genomes like that of *Brachypodium distachyon* or *Arabidopsis thaliana* contain around 20 – 30% of TEs, while species with larger genomes like *Zea mays* (maize) and *Hordeum vulgare* (barley) bear up to 85% of TEs (The Arabidopsis Genome Initiative, 2000; Wicker et al., 2005; Schnable et al., 2009; Tenaillon et al., 2010; The International *Brachypodium* Initative, 2010).

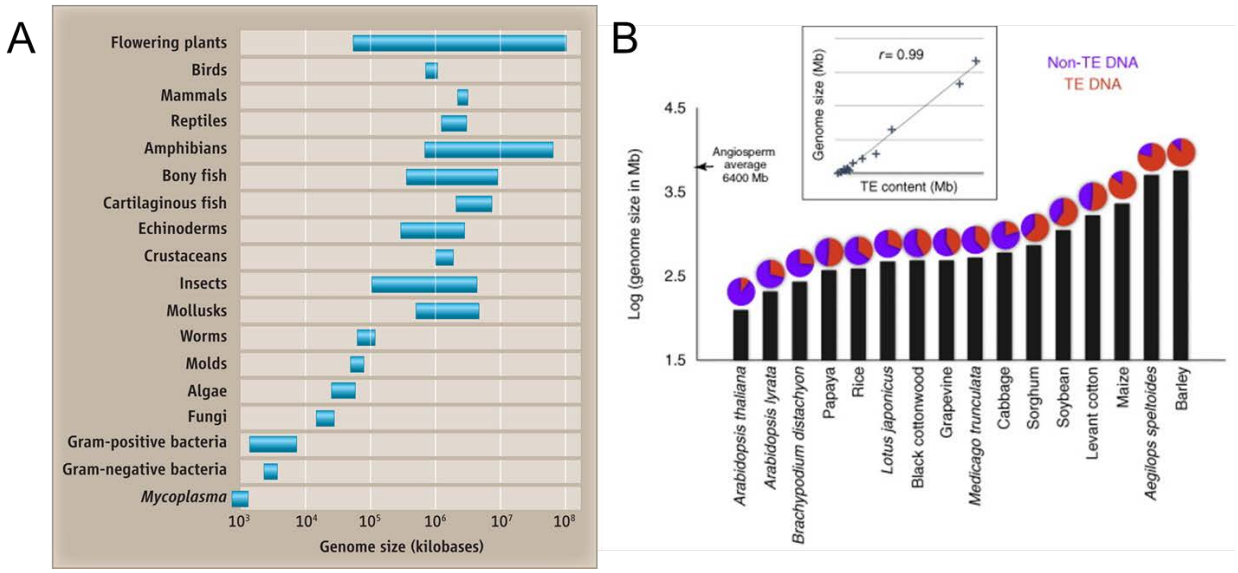

Figure 1-5: Genome sizes (A) Genome sizes estimated by the C-value (modified after (Fedoroff, 2012)). (B) Different genome sizes of angiosperms (bars) and their TE and non TE constitution (pie-charts). The plot shows the positive correlation of increasing genome size and TE content (modified after (Tenaillon et al., 2010))

Comparative genomic studies in e.g. maize, *Gossypium* spp. and *Oryza australiensis* indicate 100% increase of the genome size due to TE proliferation in short evolutionary times (SanMiguel et al., 1998; Hawkins et al., 2006; Piegu et al., 2006; Tenaillon et al., 2010). Majority of the genome size differences is attributable to Class I TEs, due to their “copy and paste” proliferation mechanism that produces new copies each amplification round (SanMiguel et al., 1998; Hawkins et al., 2006; Piegu et al., 2006; Wicker et al., 2007). LTR retrotransposons constitute about 90% of the TEs in species like maize and rice (Mao et al., 2000; Wei et al., 2009; Tenaillon et al.), 42% in soybean (Du et al., 2010) and 55 % in sorghum (Peterson et al., 2002), this makes them most abundant in plants and major contributors to the C-value paradox (Tenaillon et al., 2010).

1.2.1 Transposable elements in *Brassicaceae*

With rapid increase of genomic data owing to the boost of next generation sequencing methods, a large dataset for comparative studies became publicly available. This led to new insights on TE abundance, composition and evolution among closely related species. For instance, the release of the *A. lyrata* genome revealed significant differences in TE abundance and structure in contrast to that of the close relative *A. thaliana. A. lyrata* has a larger genome than *A. thaliana (125 Mb versus 207 Mb, respectively),* by a larger number of TEs within the genome, which fit to the proposed correlation between genome size and TE abundance (Tenaillon et al., 2010). However, TEs in *A. lyrata* appeared relatively younger and physically closer to genes (Hu et al., 2011). In addition, partially transcriptional activity of TEs was shown in *A. lyrata*, which was proposed to be due to more multiple mapping than single mapping siRNAs (Hollister et al., 2011; He et al., 2012). Multiple mapping siRNA are less efficient in silencing due to dilution by targeting and pertain to recent TE insertions (Hollister et al., 2011). Recent release of the *Arabis alpina* genome showed TE activity of *Gypsy* LTR retrotransposons, that led to a transformation of euchromatic gene clusters into repeat-rich pericentromeric regions (Willing et al., 2015). This lies in contrast to investigations on self-compatible *Capsella rubella* and self-incompatible *Capsella grandiflora*, which showed high similarity in TE abundance and age to that of *A. thaliana* without any significant difference due to the mating system (Slotte et al., 2013). These recent findings indicated large difference in the

control and regulation of TEs, leading to a different TE abundance within the genomes. Further it was shown that TE abundance seems to be independent from the mating system. However, the reason for these genome-wide differences in TE abundance within close related species remained unclear. Difference in the regulation and control of TEs within the different species was assumed and needs to be investigated under natural and challenging conditions.

1.3 Epigenetic control of plant transposable elements

Transposon activity can lead to genetic mutations such as e.g. loss-of-function gene mutations, duplications and changes in genome structure (Tenaillon et al., 2010). In order to protect genome stability, TE activity needs to be inhibited, which is established and maintained over multiple rounds of cell divisions (mitotically) and generations (meiotically) (Lisch, 2009). Transcriptional inactivation of TEs in plants is established and maintained epigenetically. The term epigenetic describes heritable changes in expression without changes in the DNA sequence (Chandler et al., 2004). Transposon activity is suppressed by transcriptional gene silencing (TGS). To ensure stable TGS, an active transposon that is transcribed by RNA polymerase II is originally a target of post-transcriptional gene silencing (PTGS) (Matzke et al., 2014).
TGS represses transposable elements by epigenetic marks as DNA methylation (5-methyl-cytosine; 5mC), high nucleosome density which can be modified by repressive epigenetic marks such as di-methylation of lysine 9 at histone H3 (H3K9me2). DNA methylation can occur in three different sequence contexts: CG, CHG and CHH, where H is A, T or C. *De novo* methylation in plants is established by the RNA-directed DNA methylation (RdDM) pathway in all three contexts at small interfering RNA (siRNA)-DNA homology sites (Matzke et al., 2009). Plant specific RNA Polymerase IV transcribes a TE locus which produced RNA is subsequently copied by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) to produce double-stranded RNA (dsRNA). Developed double stranded RNA (dsRNA) is cleaved by DICER-LIKE 3 (DCL3) into 24-nucleotide (24-nt) siRNA and subsequently methylated by HUA-ENHANCER 1 (HEN1). Silencing effector ARGONAUTE 4 (AGO4), or its closely related family members AGO6 or AGO9, bind methylated 24-nt siRNA and interact with WG/GW-motif of KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) and the C-terminal domain of NRPE1, the largest subunit of RNA

Polymerase V. This complex facilitates *de novo* methylation of the siRNA targeted site by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Figure 1-5). Maintenance of DNA methylation is established by METHYLTRANSFERASE 1 (MET1) in CG context and CHROMOMETHYLASE 3 (CMT3) in CHG context, copying particular methylation pattern to the daughter strand (Kankel et al., 2003; Fedoroff, 2012; Pikaard et al., 2014).

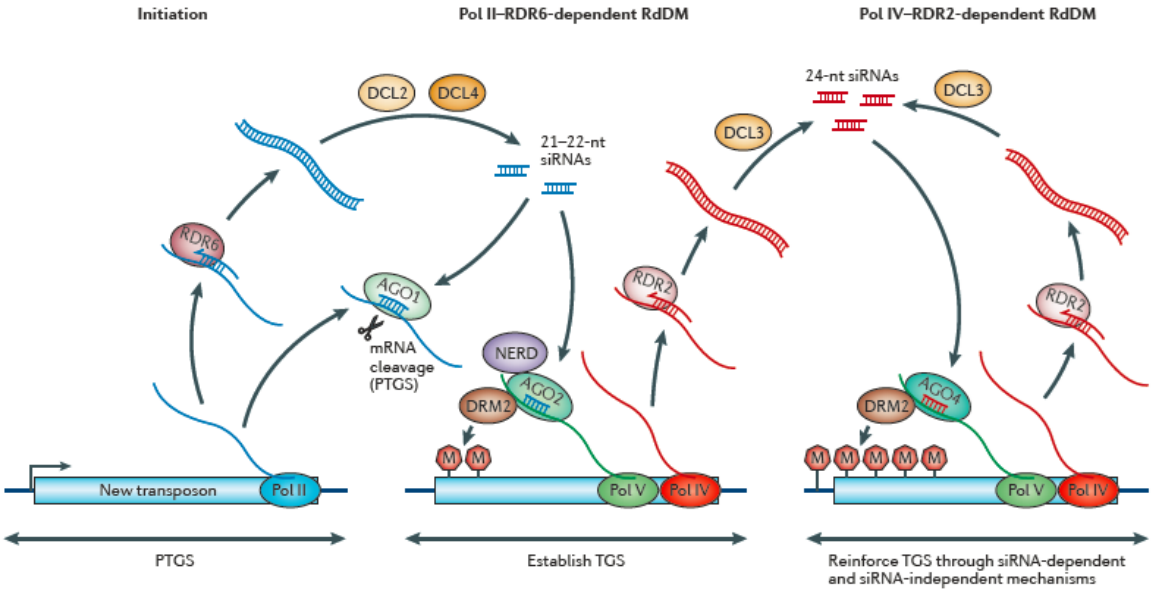

Figure 1-6: Non-canonical Polymerase-II-RDR6-dependent RNA-directed DNA methylation-pathway.

Post transcriptional gene silencing (PTGS) of TEs is carried out on RNA polymerase II (Pol II) transcribed RNA e. g. initially active TEs. These transcripts are copied by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) into dsRNA which gets subsequently processed by DICER-LIKE 2 (DCL2) and DCL4 into 21and 22 nucleotide (nt) small interfering RNA (siRNA). These siRNA are loaded into AGO1 and guides cleavage of transposon transcripts. In deviation with canonical TGS by RdDM, 21 and 22 nt siRNA can trigger *de novo* DNA methylation at a low level in a manner that is dependent Pol V scaffold transcripts, DRM2 and AGO2, which interacts with NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION (NERD) through its AGO hook motif (Figure: 1-6; (Matzke et al., 2014)).

Recent study on *COPIA93* family element *EVADÉ* (*EVD*) suggests a shift from PTGS to TGS by particular number of genomic copies. This was indicated by an accumulation of LTR mapping siRNA and decreasing transcript level of *EVD* after reaching ~40 genomic copies (Mari-Ordonez et al., 2013).

In addition to DNA methylation, transposons are regulated by chromatin accessibility. Therefore, TEs are frequently located in densely packed chromatin (heterochromatin) regions. Here, chromatin is packed into structures called nucleosomes, which consist of the DNA, snRNA and histones. In a nucleosome ~147 bp DNA is wrapped around a histone. Histones are octamers consisting of two molecules of H2A, H2B, H3 and H4, respectively. The formation of this structure ultimately results in a dense packing of the chromatin making it less accessible. Thereby, activity of TEs gets suppressed to prevent relocation and amplification within the genome.

1.4 Transposons in gene function and regulation

TEs are generally considered as selfish DNA (Doolittle et al., 1980; Orgel et al., 1980). The self-promoting mobility and amplification in host genomes may lead to mutations and gene regulatory patterns (Hua-Van et al., 2011). As mentioned before, retrotransposons are proposed to have the largest impact on the genomes. Retrotransposition has been reported to change gene regulation and function ((Lisch, 2013a; Lisch, 2013b)). One of the most striking TE induced changes are the inactivation of genes due to TE insertions. In *Setaria italica* (foxtail millet), the 431 analyzed landraces are either waxy or “sticky” due to differences in the amylose content. This phenotype is induced by null-mutations in the starch synthase gene *GBSS1* induced by TE insertions (Kawase et al., 2005; Lisch, 2013a). In addition, TE insertions in promotor regions or upstream of the transcriptional start site (TSS) of a gene can influence gene expression and regulation. One of the best studied examples in plants is the *FLOWERING WAGENINGEN (FWA)* gene in *Arabidopsis thaliana*. *FWA* is a maternally imprinted gene, which contains a *SINE* element in its promoter region which ensures epigenetic silencing of this gene in vegetative tissues (Soppe et al., 2000). In addition, natural variation in the *FWA* promotor in different *Arabdidopsis* species like duplication, triplication and quadruplication of this *SINE* element leads to expression of *FWA* also in vegetative tissues in *Arabidopsis arenosa*, and specific accessions of *A. lyrata* and *A. halleri* (Fujimoto et al., 2008). In contrast, stress induced expression of TEs leading to new insertions adjacent to genes has shown that neighboring genes become stress responsive as well. This was observed in *Citrus sinensis* varieties (1.4), but also in post heat stress progenies of *A. thaliana* RdDM-mutant (Ito et al., 2011; Butelli et al., 2012). *ONSEN*

transposition adjacent to *ACTIVATED DISEASE RESISTANCE GENE 1* (*ADR1*) promotes its expression under heat stress conditions (Ito et al., 2011). In contrast, gene regulation can also be altered because of TE mediated gene movement or retrotransposition to another chromosomal locus (Lisch, 2013a). Gene movement was observed, for instance, in *Solanum* varieties. Variation of the *SUN* locus is responsible for differences in fruit shape in *Solanum lycopersicum* varieties and its wild relative *Solanum pimpinellifolium (van der Knaap et al., 2004)*. *IQ domain 12* (*IQ12*) is the key gene at the *SUN* locus, its retrotransposition to another gene led to gene expression regulation by the new promotor (Xiao et al., 2008). Retrotransposition happens whenever translated proteins of a retrotransposon targets gene transcripts than its TE origin (Kaessmann et al., 2009). Recent publication indicated 251 retrogenes in *A. thaliana* corresponding to 1% protein-coding genes. While 25% are cotranscribed with their parents, 3% are head-to-head oriented neighbors, suggesting 72% of *Arabidopsis* retrogenes being regulated by novel promotors (Abdelsamad et al., 2014).

Impact of TEs on chromosomal structure is well documented in maize (Lisch, 2013a). Transposition of the *AC* elements in maize led to deletions, inversions, translocations and other chromosomal rearrangements (Yu et al., 2011).

Activity is only one force leading to TE-driven genome evolution (Tenaillon et al., 2010). Illegitimate recombination and unequal intra-strand homologous recombination eliminate TEs from its host genome (Le et al., 2000; Devos et al., 2002; Pereira, 2004; Sabot et al., 2011), which can be a result of several factors like demographic history and propagation systems (Wright et al., 2000; Morgan, 2001; Lockton et al., 2008).

1.5 Abiotic stress-mediated TE expression

There is a growing body of literature indicating that TEs can be activated in response to biotic and abiotic stresses e.g. salt (Naito et al., 2009), wounding (Mhiri et al., 1997), bacteria (Grandbastien et al., 2005), cold (Grandbastien et al., 2005; Naito et al., 2009) and heat (Pecinka et al., 2010). One of the most famous examples was observed in the development of blood oranges (*Citrus sinensis*). Insertion of the *COPIA*-family TE *Rider,* upstream of the *Ruby* gene, a transcription factor of anthocyanin production, resulted into cold-induced expression of this gene in the

fruit. In contrast, a variety which lacks of that insertion shows limited expression of *Ruby* and thus anthocyanin poor phenotype (Butelli et al., 2012). Interestingly, Ivashuta et al. (2002) observed novel TE called *MCIRE* in *Medicago sativa* whose expression is induced by a low temperature responsive element (LTRE) in the 5´LTR. Similar observations were monitored in *TRANSPOSABLE ELEMENT OF NICOTIANA TABACUM 1* (*Tnt1*) elements that showed transcriptional activity to plant defense reactions, wounding and freezing (Beguiristain et al, 2001; Casacuberta et al, 1997; Mhiri et al, 1997) or at light and salinity activated *TRASPOSABLE ELEMENT OF TRITICUM DURUM* (*Tdt1*) (Woodrow et al., 2010; Woodrow et al., 2011). Particular motifs triggering the expression of these LTR retrotransposons reside in their 5´LTR and are similar to regulatory motifs in particular stress responsive genes (reviewed in (Casacuberta et al., 2013)).

One of the best studied examples of stress induced expression by a regulatory motif in the 5´LTR is the heat responsivity of *COPIA78* TEs in *A. thaliana* accession Col-0 (Pecinka et al., 2010; Cavrak et al., 2014) named *ONSEN (Ito et al., 2011)*. *ONSENs* are autonomous *Ty1/COPIA* LTR-retrotransposons within the *COPIA78* family which are characterized by transcriptional up-regulation under heat stress conditions and slow transcript decrease during recovery of plants at standard growing conditions (Pecinka et al., 2010). *ONSEN* transcription was shown to be mediated by the *HEAT SHOCK FACTOR A1* (*HSFA1*)-pathway. It was shown that the triple mutants of the *HSFA1* genes a/b and d lead to altered heat induced transcription of *ONSEN*. Further, HSFA1 paralogs needs to form homo- or heterotrimers in order to activate downstream targets (Yoshida et al., 2011; Cavrak et al., 2014). One of these targets is the *HEAT SHOCK FACTOR A2 (HSFA2)*, which is highly expressed during heat stress in *Arabidopsis (Schramm et al., 2006)*. It was described that *HSFA2* binds to a heat responsive element (HRE) sequence motif in the *ONSEN*-LTR consisting of two palindromic repeats (nTTCnnGAAn (Cavrak et al., 2014)). New *ONSEN* insertions were only found in progenies of heat stressed *RDR2 and NRPD2A,* the second largest subunit of Pol IV*,* mutants (Ito et al., 2011; Matsunaga et al., 2012). However, retrotransposition occurs during flower development and before gametogenesis. In progenies of heat stressed siRNA deficient mutants, new *ONSEN* copies adjacent to genes were shown to promote heat responsivity of these (Ito et al., 2011).

Recent study indicated presence of *ONSEN* elements also in other *Brassicaceae* species e.g. *Brassica rapa* and *A. lyrata*, however it also indicated *ONSEN* absence in *Capsella* and *Cardamine hirsuta.* In addition, heat induced transcriptional activity of *ONSEN* in other *Brassicaceae* was shown, indicating evolutionary conserved heat responsivity. However, the mechanism leading to this conserved *ONSEN* heat responsiveness remains unknown. It is also still cryptic if heat responsiveness is conserved within one particular class or superfamily e.g. COPIA LTR-retrotransposons and if it evolved repeatedly within other TEs.

1.6 Aims of the PhD thesis

Eukaryotic genomes consist of a large number of TEs. The selfish and parasitic behavior of TEs has been investigated extensively. However, there is growing evidence that TEs contribute to genome evolution (The Arabidopsis Genome Initiative, 2000; Tenaillon et al., 2010; Hollister et al., 2011; Hu et al., 2011; Ito et al., 2013; Lisch, 2013a; Slotte et al., 2013; Willing et al., 2015). TEs are strongly suppressed by epigenetic modifications. Therefore it is unclear how TEs escape epigenetic silencing and amplify within the genome. Recent studies indicated that specific *COPIA78* TEs in *A. thaliana* named *ONSEN* become activated in response to heat stress, but remained suppressed also in mutants deficient in epigenetic silencing (Pecinka et al., 2010; Tittel-Elmer et al., 2010). *ONSEN* heat stress induced activation was shown to be mediated by specific heat shock transcription factor DNA binding motifs in their LTRs (Cavrak et al., 2014). Such transcription factor binding motifs are similar to regulatory motifs found in stress responsive genes (Casacuberta et al., 2013). In addition, presence and partially heat induced *ONSEN* transcription was shown in other Brassicaceae species (Ito et al., 2013). The aim of this study is to investigate the evolution of TE heat responsiveness in *Brassicaceae* and answer the following questions:

I. Are there novel or new common heat responsive TEs within *A. thaliana* and *A. lyrata* and do they share common heat responsive DNA motifs (HRE)*?*
II. Do other *Brassicaceae* possess those heat responsive TEs and do they contain HREs within their LTR?
III. Considering this how did TE heat response evolved within *Brassicaceae*?

2 Materials and Methods

2.1 Plant materials

Initial experiments and whole transcriptome sequencing were performed on *Arabidopsis thaliana* accession Col-0 and *A. lyrata* accession MN47.

For interspecific analysis of TE occurrence and heat induced expression, *Arabis alpina*, *Boechera stricta*, *Brassica rapa* FPsc, *Capsella rubella* and *Eutrema salsugineum* were added to analysis.

2.2 Seed sterilization and *in vitro* plant growth

Seeds were sterilized with sodium hypochlorite. Whole procedure was carried out under a sterile hood. Therefore, seeds were filled into a 1.5 ml reaction tube. Subsequently, 500 µl of 70% Ethanol were added and incubated at room temperature for 5 minutes. After that, Ethanol was removed and 500 µl of 8% sodium hypochlorite was added and incubated for a maximum of 12 minutes. Sodium hypochlorite was removed and the seeds were washed 4 times in 1 ml sterile water. After the last washing step, the seeds were resuspended 1 ml of 0.1% agarose solution.

Seeds for interspecific analysis of TE occurrence and heat induced expression were sterilized (2.2.) and sown on ½ MS-media with 50 µM Gibberellic acid 4+7 (Duchefa Biochemie/Netherlands) and stratified. Stratification was conducted in a dark room at 4°C. *A. thaliana* Col-0, *A. lyrata* MN47, *A. alpina*, *B. rapa* and *C. rubella*, were stratified for 1 week, *B. antipoda* and *B. stricta* for 2 weeks and *E. salsugineum* for 3 weeks, to achieve successful germination and uniform growth rates. Subsequently, seeds were transferred to a growth chamber (Percival CU-36l5; Percival Scientific / USA) where they were germinated and grown under short day conditions (8h 21°C light / 16h 18°C dark) 60 – 65% and a 150 µmol m^{-2} s^{-1} light intensity till seedling stage. Seedlings were transferred to single pots with ½ MS-media and grown under short day conditions.

2.3 Plant growth

For initial experiments and whole transcriptome analysis, seeds of *A. thaliana* Col-0 and *A. lyrata* MN47 were sown on moist soil and stratified for 7 days at 4°C in a dark room, to achieve uniform germination. Subsequently, seeds were transferred to a growth chamber (Percival AR-95L3; Percival Scientific / USA) and were germinated and grown under long day conditions (16h 21°C light / 8h 16°C dark) with a relative humidity of 60 – 65% and a 150 µmol m^{-2} s^{-1} light intensity. After two weeks, plants were singled to one plant per pot to prevent plants from additional stresses by space or nutrition competition and to obtain independent biological replicates.

2.4 Heat stress and recovery treatments

Heat stress treatments for differential TE expression and whole transcriptome analysis in *A. thaliana* Col-0 and *A. lyrata* MN47 were conducted at 37°C in a preheated growth chamber. Therefore, plants at 5 leaves rosette stage were transferred to the 37°C chamber in a tray closed with a lid containing small holes at the top to assure air flow, but preventing plant drying due to transpiration. The samples were collected after 1, 3, 6, 9, 12 and 24 hours 37°C. For recovery, 6, 12 and 24h heat-stressed plants were transferred back to normal growth conditions and harvested after 48 h.

Interspecific comparison of heat induced transcription of *ONSEN*, *HATE*, *COPIA20* and *COPIA37* was performed in a growth chamber (Percival CU-36l5; Percival Scientific / USA) preheated to 37°C. Therefore, plants grown on ½ MS-media until 5 leaves rosette stage, were transferred to the 37°C chamber. Samples were harvested after 6 and 12 hours at 37°C.

2.5 Molecular biological methods

All standard and not further explained methods were conducted as described in Sambrook (2001) and Weigel et al. (2002).

2.5.1 Chemicals

All chemicals, which are not specifically mentioned, were purchased from the following companies: Applichem (Darmstadt/Germany), Bio-Budget Technologies

(Krefeld/Germany), Bio-Rad (München/Germany), Carl-Roth (Karlsruhe/Germany), Duchefa Biochemie (Haarlem/Netherlands), Fermentas/ThermoScientific (St. Leon-Rot/Germany), Life Technologies (Karlsruhe/Germany), Merck (Darmstadt/Germany), New England Biolabs (Frankfurt am Main/Germany), peqLab (Erlangen/Germany), Promega (Mannheim/Germany), Roche (Basel/Switzerland), Sigma-Aldrich (Steinheim/Germany), VWR International (Darmstadt/Germany).

2.5.2 DNA-isolation

For DNA isolation, 2 – 3 leaves were collected and frozen in liquid nitrogen. DNA isolation for standard PCR reactions e.g. control reaction in cDNA validation, was performed with the Nucleon PhytoPure gDNA Kit (GEHealthcare; München/Germany) according to manufacturer´s instructions.

2.5.3 RNA-isolation

Total RNA was isolated from approximately 2 - 3 rosette leaves frozen in liquid nitrogen, by using the RNeasy Plant Mini Kit (Qiagen, Hilden/Germany) with an additional on-column DNase I digestion (Roche, Basel/Switzerland).

2.5.4 Oligonucleotides

All Oligonucleotides used in this study were synthesized by Metabion (Martinsried/Germany) or Life Technologies (Karlsruhe/Germany). Oligonucleotides used for transcription analysis were designed manually by using Oligo Calculator version 3.07 (http://www.metabion.com/biocalc/index.html) or Primer3 (Koressaar et al., 2007; Untergasser et al., 2012). The used Primers are enlisted in the appendix (Table A1).

2.5.5 Polymerase chain reaction (PCR)

Standard PCR reactions e.g. for testing cDNA purity were processed with Taq DNA Polymerase (NEB; Frankfurt am Main/Germany).

2.5.6 Transcription analysis

RNA concentrations were quantified with the NanoDrop ND-1000 spectral photometer (peqLab, Erlangen/Germany). An amount of 1 µg of total RNA was reverse transcribed into complementary DNA (cDNA) with the First Strand cDNA Synthesis Kit using random hexamer primers (Fermentas/ThermoScientific, St.Leon-Rot/Germany) and analyzed by quantitative reverse Transcriptase PCR (qRT-PCR) with the SensiMix SYBR & Flourescein Kit (Bioline; Berlin/Germany). Transcription levels of TEs were estimated by the standard curve method and normalized to the heat stress stable reference genes *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C-2* (*GAPC-2*; AT1G13440) or *UBIQUITIN-CONJUGATING ENZYME 28* (*UBC28*).

2.5.7 Whole transcriptome sequencing

For whole transcriptome sequencing, total RNA was extracted as described (2.2.3). RNA concentrations were quantified by spectrophotometry using the Qubit RNA HS Assay kit and the Qubit Fluorometer (Life Technologies, Karlsruhe/Germany). RNA-Integrity Numbers (RIN) were determined on a Bioanalyzer assay using the Agilent RNA 6000 Nano Kit (Böblingen/Germany). Samples with a RIN between 8 and 10 were used for library construction. RNA Libraries were made using Illumina TruSeq RNA Sample Prep Kit (San Diego/USA) following manufacturer's instructions. Subsequently, library concentrations were measured with the Qubit dsDNA HS Assay Kit on the Qubit Fluorometer and its insert size and integrity analyzed on a Bioanalyzer using the Agilent DNA 1000 Kit (Böblingen/Germany).

High throughput sequencing was performed on an Illumina HiSeq2500 sequencer with a requested sequencing depth of 18.7 million 100bp single end reads per library at the Max Planck Genome Center (Cologne/Germany).

2.6 Bioinformatic methods

2.6.1 Statistical analysis

Statistical analyses were performed with “R” (http://www.r-project.org) and the appropriate packages. Charts were constructed within “R” or Microsoft Excel. Chi-square tests with Yates-correction were performed using Graph-Pad Software online tool (http://graphpad.com/quickcalcs/contingency1.cfm).

2.6.2 Whole transcriptome analysis

Obtained RNA-sequencing raw reads where quality controlled using fastqc v0.10.1 (Andrews). Subsequently, adapter sequences and low quality bases were trimmed with the FASTX-toolkit (García-Alcalde et al., 2012) using standard parameters. The libraries with sufficient quality, were mapped to corresponding reference genome *A. thaliana* Col-0 TAIR10 (Lamesch et al., 2011) or *A. lyrata* genome assembly v1.0 (Hu et al., 2011) using bowtie2 and TopHat2 with default parameters (Langmead et al., 2012; Kim et al., 2013). After mapping, read counts per position were estimated using species specific TE annotation developed within Pecinka lab by A. Abdelsamad (unpublished). Differential expression of TEs between different conditions per species where calculated in R software (R Core Team, 2013), with the DEseq package (Anders et al., 2010).

2.6.3 Detection and reconstruction of long terminal repeats (LTRs)

Detection of LTRs from *A. thaliana* and *A. lyrata* TEs were performed by using LTR_Finder (Xu et al., 2007) or by pairwise nucleotide sequence alignments using Multiple Sequence Comparison by Log-Expectations (MUSCLE; (Edgar, 2004)) online tool (http://www.ebi.ac.uk/Tools/msa/muscle/). Detection of LTRs in *A. thaliana* using LTR_Finder were conducted with default Parameters using *A. thaliana* (Feb. 2004) tRNA database to predict Primer binding site (PBS) for reverse Transcriptase initiation provided within the tool. For *A. lyrata ONSEN*-LTR detection using LTR_Finder were used with decreased output score threshold to 3.0 (default = 6.0) and *A. thaliana* (Feb. 2004) database for PBS prediction. However, in cases where LTR detection by LTR_Finder was impossible, pairwise nucleotide sequence

alignments were performed in MUSCLE (Edgar, 2004) online tool with default parameters.

TEs and LTRs in species without public available TE annotation were identified by using the Basic Local Alignment Search Tool (BLAST, (Altschul et al., 1990; Goujon et al., 2010; McWilliam et al., 2013)) on Phytozome v10 (Goodstein et al., 2012). Therefore, *A. thaliana ONSEN, A. lyrata ONSEN, HATE, COPIA20* and *COPIA37* sequences and identified LTRs were aligned versus whole genome sequence of *B. stricta* v1.2 (DOE-JGI, http:://www.phytozome.net/bstricta), *B. rapa* FPsc v1.3 (DOE-JGI, http:://www.phytozome.net/BrapaFPsc), *Capsella grandiflora, C. rubella,* (Slotte et al., 2013)*, E. salsugineum* (Yang et al., 2013). Detection of TEs in *A. alpina* was performed by BLAST against a self-created nucleotide database of *A. alpina* genome fasta-file (Willing et al., 2015) via BioEdit software (Hall, 1999). TEs in *B. antipoda* were identified by using BLAST within the CLC workbench (http://www.clcbio.com) against a self-created nucleotide database build from *B. antipoda* contigs (Vu, Finke, Pecinka; unpublished data). Sequence information and genomic location of BLAST results were extracted and used to reconstruct LTR sequence by a hierarchical cluster analysis (Corpet, 1988) at MultAlin (http://multalin.toulouse.inra.fr/multalin/) with default parameters of DNA comparison table “DNA-5-0”. Subsequently, profile hidden Markov model search was used to verify results by using nHMMer within HMMMer3 with standard parameters (http://hmmer.janelia.org/; (Mistry et al., 2013; Wheeler et al., 2013))

2.6.4 Re-annotation of *HEAT ACTIVATED TRANSPOSABLE ELEMENT* (*HATE*)

HATE in *A. lyrata* was re-annotated as proposed in Wicker et al. (2007). First, sequences were analyzed by a BLASTN against the TE database of RepeatMasker using “cross_match” search engine. Due to their low sequence similarity (< 70%) a BLASTX was performed using NCBI BLASTX online tool on *HATE* sequences, to obtain information about the constitution of the GAG and POL Domains. This indicated *HATE* as being *Ty1/COPIA*-like elements. Additional information about LTR´s, PBS, PPT and TSDs was obtained by using LTR-Finder online tool.

2.6.5 DNA sequence analysis

All *in-silico* sequence analysis, e.g. detection of heat responsive elements, was conducted in SeqBuilder from the DNASTAR Lasergene 9 Core Suite (DNASTAR Inc., Madison/USA).

2.6.6 Phylogenetic reconstruction

Multiple sequence alignments were carried out using MUSCLE within the EMBOSS-Package (Rice et al., 2000) on a Linux cluster or by using the MUSCLE online tool (Larkin et al., 2007). All multiple sequence alignments were conducted by default settings unless specified otherwise. Alignments were saved in clustal or fasta file-formats and loaded into the Program MEGA6 (Tamura et al., 2013) for further analysis. Within MEGA6, the evolutionary history was inferred using the Neighbor-Joining method (Saitou et al., 1987). The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the analyzed taxa (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980). TE age were estimated by pairwise base-pair distance of 3´and 5´LTR in MEGA6.

2.6.7 Phylogenetic network construction

Phylogenetic networks were constructed using Neighbor-Net (Bryant et al., 2004) within the splitstree 4.0 package. (Huson et al., 2006). The phylogenetic distances were calculated by using UncorrectedP-distance that computes the proportions of positions in which two sequences differ within the splitstree 4.0 package.

2.6.8 Phylogenetic shadowing

Phylogenetic shadowing and motif conservation was conducted using mVISTA (Mayor et al., 2000; Frazer et al., 2004). As a prerequisite, LTR consensus sequences of the different species were created in the program BioEdit (Hall, 1999) allowing ambiguity code except of gaps for low conserved positions. The consensus sequences were loaded into mVista web application and aligned using AVID (Bray et al., 2003). Conservation of functional regions was calculated within a sliding

conservation window of 8 base pairs and a minimal conservation width of 10 base pairs within the Vista Browser java applet. Conserved identity threshold was empirically set to 70%.

3 Results

3.1 Repetitive DNA elements in *A. lyrata* MN47 and *A. thaliana* Col-0

In order to analyze the evolution of TE heat responsiveness, TEs that are transcriptionally active in response to heat within other species needed to be identified. Here we used *A. lyrata*, a species closely related to *A. thaliana* and diverged from a common ancestor approximately 10 million years ago (Wright et al., 2002; Beilstein et al., 2010; Ossowski et al., 2010; Hu et al., 2011). In spite of the close relatedness, analysis of the recently assembled *A. lyrata* genome sequence revealed specific differences in genome size and composition (Figure 3-1; (Hu et al., 2011)). While *A. thaliana* has a reduced chromosome number to n = 5 and its genome is only approximately 135 Mbp, *A. lyrata* possesses the ancestral *Brassicaceae* chromosome count of n = 8 and has considerably larger genome with 207 Mbp (Lysak et al., 2006; Hu et al., 2011). However, both genomes contain similar number of genes (32,041 versus 33,298, respectively) and 90 % of genic regions are syntenic between both genomes (Hu et al., 2011; Rawat et al., 2015). In contrast, major differences exist in the copy number of TEs, where *A. lyrata* contains 53,090 elements compared to 17,009 elements in *A. thaliana*. Although this explains the genome size differences between both species to some extent, even more important factor driving genome reduction in *A. thaliana* genome seems to be deletion-biased repair of DNA strand breaks, resulting in hundreds of thousands small deletions (Hu et al., 2011).

A

	 *A. thaliana* Col-0	 *A. lyrata* MN47
Genome size (Mb) [1]	125	207
Chromosomes [1]	5	8
Genes [2]	32,041	33,298
TEs (n) [3]	17,009	53,090
TE families (n) [3]	364	376

B

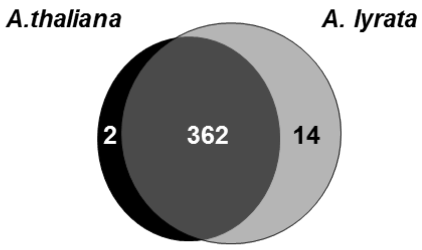

Figure 3-1: *A. thaliana* and *A. lyrata* genome features. (A) Genome structure and composition overview. (B) Venn diagram of TE families in both species. [1]Hu et al. (2011) [2](Rawat et al., 2015) [3]This study.

While there is available robust annotation of protein coding genes for both species (The Arabidopsis Genome Initiative, 2000; Rawat et al., 2015), their TE annotation is much less developed. Therefore, custom-made TE annotation files were developed using identical settings by Dr. A. Abdelsamad within the Pecinka Lab (unpublished data). Subsequently, TE annotation files were inspected for unmerged annotation units of neighboring elements using BLASTX (translated DNA versus Protein database), LTR_finder and multiple pairwise alignments with corresponding TE families (Altschul et al., 1990; Larkin et al., 2007; Xu et al., 2007; Goujon et al., 2010; McWilliam et al., 2013), and the errors were corrected manually. This resulted in annotation of 53,090 TEs in 376 TE families occupying 17.5% of the *A. lyrata* genome. The *A. thaliana* custom-made TE annotation contains 17,009 elements in 364 families, which cover 12.8% of the genome. This created a basic dataset for comparative studies.

Comparison of TE populations revealed that the most abundant order were the DNA transposons representing 42% and 35% of annotated TEs in *A. thaliana* and *A. lyrata*, respectively (Figure 3-2). The second most abundant order was represented by LTR-retrotransposons constituting 30 % of all TEs in *A. thaliana* and 27 % in *A. lyrata*. In contrast, *Helitrons* constituted only 16 % of the *A. thaliana* and 21 % of the

A. lyrata TE set. Hence, both species differ in TE copy numbers but the overall composition of their TEs populations is relatively similar.

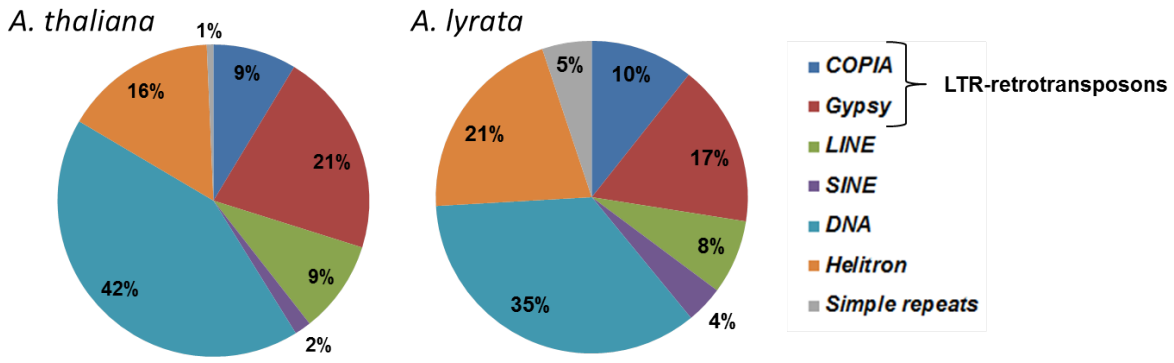

Figure 3-2: Genome-wide TE spectrum in *A. thaliana* (n = 17,009) and *A. lyrata* (n = 53,090).

3.2 Comparative analysis of heat-induced TEs in *A. lyrata* and *A. thaliana*

To identify common or novel heat responsive TEs, we performed genome-wide transcriptome analysis of heat- and mock-treated plants.

3.2.1 Establishing the effective heat stress regime

Heat exposure leading to significant transcriptional activation of TEs in both species needed to be established. Previously, the activation of TEs in *A. thaliana* was induced by long heat treatment (24 to 30 h) of in vitro grown plants (Pecinka et al., 2010; Ito et al., 2011; Ito et al., 2013). In order to perform the treatment under more natural conditions, we grew plants on soil and aimed at the shortest possible heat stress duration to reduce negative growth effects (Larkindale et al., 2002; Rizhsky et al., 2004; Pecinka et al., 2010). *A. thaliana* Col-0 and *A. lyrata* MN47 plants at approximately 5 leaves rosette stage were exposed to 37°C and aerial tissue samples were harvested after 0, 1, 3, 6, 9 and 12 h (Figure 3-3). Subsequently, transcription of *ONSEN*s per treatment relative to mock was measured by reverse transcription quantitative-PCR (RT-qPCR). Here, *A. thaliana ONSEN* 3 (*AtONSEN3*; AT5G13205), which showed the strongest response in ATH1 microarray experiment (Pecinka et al., 2010), was used as a marker to determine mRNA levels. In *A. lyrata*, multiple pairwise alignments indicated *COPIA78 (AlONSEN15)* on scaffold 247

between position 3,711 and 8,654 as the closest homolog of *AtONSEN3* (Appendix Table: A2).

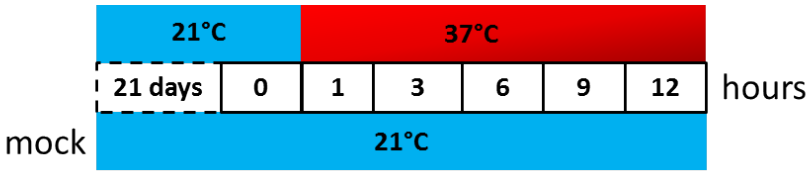

Figure 3-3: Time course that were used to detect lowest dose of heat exposition leading to transcriptional activity of *ONSEN*. Upper blocks indicate treatment for heat stress samples, lower block treatment for mock samples. Green blocks indicate 21°C growing conditions, red block indicates 37°C heat regime. Middle squares indicate sampling times, while dashed lined square indicates pre-stress plant growth.

There was similarly higher amount of *ONSEN* mRNA in both species after long heat treatment. However, accumulation of *AlONSEN* mRNA was delayed compared to *AtONSEN.* The first time point with a significantly increased relative amount of *AlONSEN* transcript (t-test; $\alpha = 0.05$) was after 6 h at 37°C. This indicated 6 h as the lowest dose necessary for significant transcriptional activation of strongly heat-responsive TEs in *A. lyrata*.

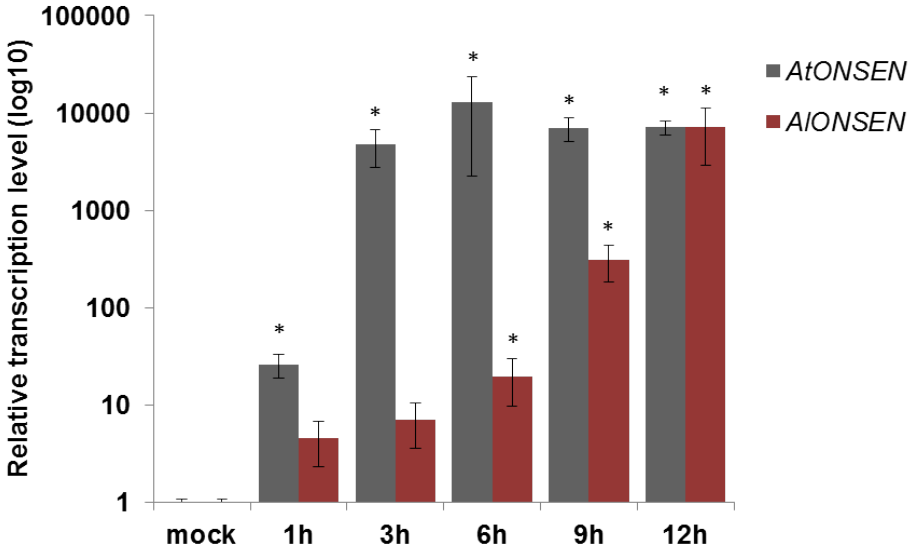

Figure 3-4: Transcript level of *ONSEN* relative to mock in *A. thaliana* (*AtONSEN*) and *A. lyrata* (*AlONSEN*) during a time course. *ONSEN* mRNA amounts were normalised to that of *UBC28* gene. Error bars are standard errors from two biological replicates. Statistical significance is indicated by asterisks (t-test; $\alpha = 0.05$).

3.2.2 Whole transcriptome analysis

Previous genome-wide search for heat-induced TEs was conducted by microarray studies in *A. thaliana* (Pecinka et al., 2010; Tittel-Elmer et al., 2010). In order to avoid selection bias by low throughput analysis and to identify novel heat responsive TEs e.g. with low transcriptional abundance (Zhao et al., 2014), whole transcriptome analysis was performed in *A. thaliana* and *A. lyrata*.

After defining the stress conditions, whole transcriptome analysis using RNA-sequencing was performed. The *A. lyrata and A. thaliana* plants were exposed to 6 h of control 21°C (mock) and 6 h of 37°C heat treatment (3.2; Figure 3-5). Additionally, part of mock and heat-treated plants was transferred to standard growth conditions for a 48 h post-stress recovery period. Samples for RNA sequencing were harvested at given time points.

Figure 3-5: Conditions and sampling time points for whole genome transcriptome analysis. Green blocks indicate 21°C growing conditions, red block indicates 37°C heat regime. Prestress plant growth indicated by dashed square, sampling timepoints indicated by squares below treatment. Poststress 21°C for 48 hours indicates recovery samples.

Differential expression of TEs was calculated separately among two treatments. First, transcriptional change of TEs after 6 hours heat treatment relative to mock was computed. Second, differential TE expression in recovery samples relative to mock was calculated.

3.2.3 Whole transcriptome analysis revealed new and novel common heat responsive TEs in *A. lyrata and A. thaliana*

Differential expression analysis in heat relative to mock treated plants revealed 190 significantly (adjusted p-value < 0.05 in DESeq) upregulated TEs in *A. lyrata* and 132 in *A. thaliana* (Figure 3-7 A, Appendix Table A3 & 4). Thus, more LTR-retrotransposons are upregulated in response to heat than DNA transposons in *A. thaliana* (72 % versus 16 %, respectively) and *A. lyrata* (44% versus 25%, respectively). This overrepresentation of LTR-retrotransposons lies in contrast to their genome frequencies.

To compare the extent of heat responsive transcriptional activity of TEs in both species, the “fold-enrichment” of transcriptional TE representation relative to their overall genomic frequencies was calculated. This showed that *COPIA* family TEs are overrepresented in *A. lyrata* and *A. thaliana* (Figure: 3-8). The “fold enrichment” relative to the genome-wide abundance was almost equal in both species.

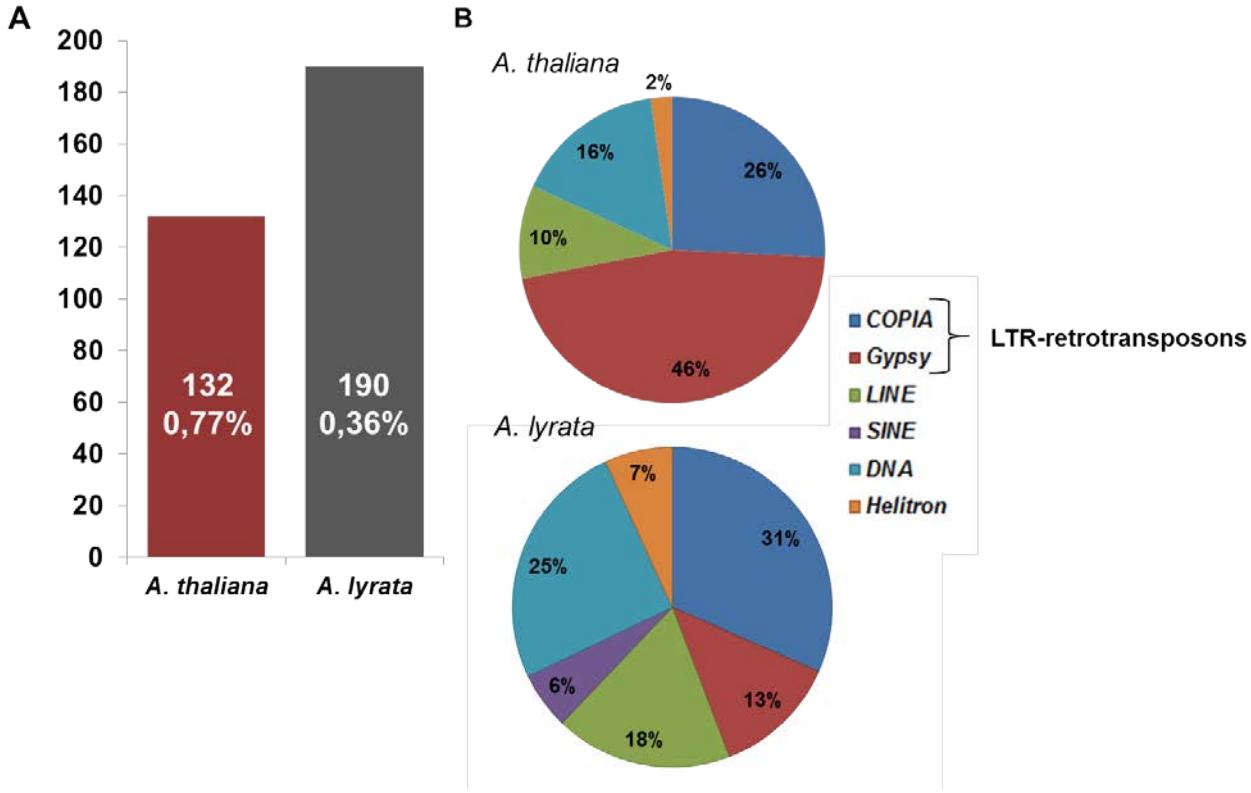

Figure 3-6: Identification of heat responsive TEs in *A. lyrata* and *A. thaliana*. (A) Significantly up-regulated TEs in *A. lyrata* and *A. thaliana*. (B) Distribution of significantly up-regulated TEs among TE superfamilies in *A. lyrata* and *A. thaliana*.

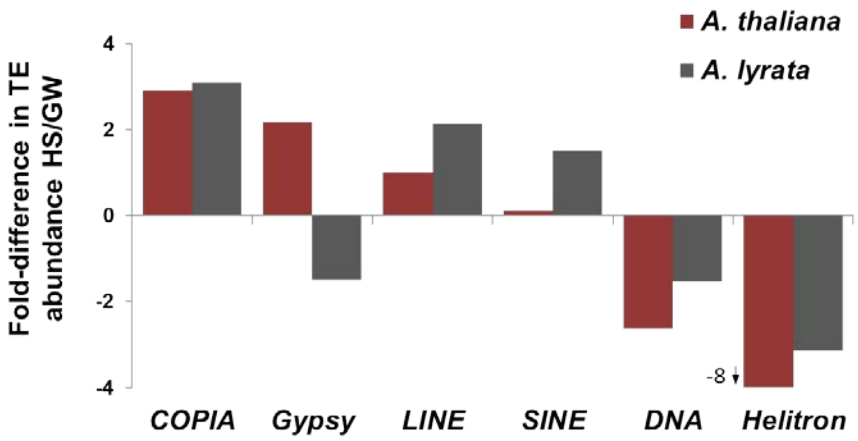

Figure 3-7: Fold-difference abundance of TE classes in *A. thaliana* and *A. lyrata* in response to heat (HS) relative to their overall genome-wide spectrum (GW).

Since *COPIA* was the one with the greatest enrichment in heat-responsive TEs, this superfamily was used as a marker group and analyzed in detail (Figure: 3-9). This revealed an overrepresentation of *ONSEN* TEs being significantly upregulated in response to heat in *A. thaliana* and *A. lyrata* (25 % versus 32 %, respectively; Figure: 3-9). In addition, significant up-regulation of *COPIA37* TEs in *A. thaliana and A. lyrata* (13 % versus 9 %, respectively) indicated a novel common heat responsive TE family (Figure 3-9). Two additional TE families were overrepresented in response to heat in *A. lyrata* - *COPIA20* and *HATE* (15% versus 10%, respectively). *HATE* is not only a heat-responsive but so far unknown family of *COPIA* TEs (3.2.4).

Summarized, these findings indicated a novel common heat responsive family within both species and hypothesizes that heat responsiveness is not exclusively conserved within *ONSEN.* In addition, putative novel heat responsive TEs were identified within *A. lyrata*, which could give new insights about the evolution of heat responsiveness in TEs.

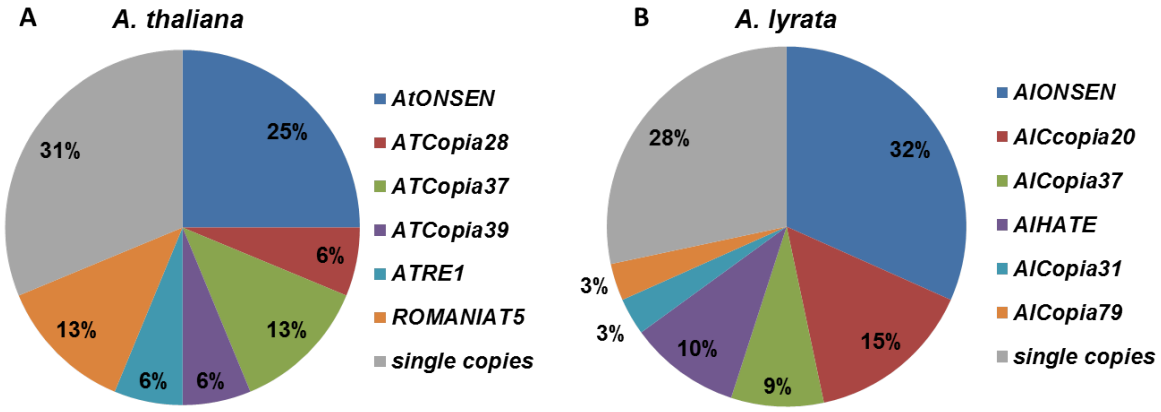

Figure 3-8: Distribution of *COPIA* families transcriptionally-induced by heat in (A) *A. thaliana* (n = 32) and in (B) *A. lyrata* (n = 60).

3.2.4 *HEAT ACTIVATED TRANSPOSABLE ELEMENT* (*HATE*) – a novel family *of Brassicaceae COPIA* elements

Transcriptome analysis revealed heat-responsiveness of *A. lyrata COPIA46* family. However, sequence analysis of these elements indicated a low query coverage and sequence similarity in comparison to *A. thaliana COPIA46* (Table: 3-1).

Table 3-1: Comparison of *HATE* with the most similar *COPIA* elements. Overview of compared *A. thaliana* and *A. lyrata* TE sequences vie multiple pairwise alignments. Values are represented as mean of pairwise sequence identities. Query sequence coverage of unequal TEs estimated between 3 -40 %. Note that no significantly similar sequences to *AlHATE* were found in *A. thaliana* Col-0 genome.

		A.thaliana		***A. lyrata***		
		ONSEN	*COPIA46*	*ONSEN*	*COPIA46*	*HATE*
A. thaliana	*ONSEN*	98 %				
	COPIA46	37 %	96 %			
A.lyrata	*ONSEN*	91 %	37 %	98 %		
	COPIA46	66 %	72 %	48 %	92 %	
	HATE	47 %	33 %	69 %	32 %	97 %

Such bias in sequence coverage and similarity among TEs of the same family needed to be observed. Therefore, we performed more detailed analysis of this element using previously proposed strategy by Wicker et al. (2007). First, sequences were BLASTN searched against a TE database in RepeatMasker (Smit et al., 2013-2015) using the cross_match search engine. The results indicated a lower (< 80%) sequence similarity of the 5´- and 3´- fragments of *COPIA78/ONSEN*-LTRs and a low (< 70 %) sequence similarity of the coding region of *COPIA46* GAG and POL Domains with large gaps between annotation units (200 – 2000 bp). The coding region represented the largest annotation unit, which apparently led to assignment of those elements as *COPIA46* in an automated annotation.

Therefore, we performed *de novo* analysis to identify the corresponding TE class and superfamily for those TEs using BLASTX (Altschul et al., 1990; Wicker et al., 2007). All these elements were unambiguously identified as *Ty1/COPIA* LTR-retrotransposons, according to their GAG- and POL-Domain composition (Figure: 3-9). Subsequently, a multiple global pairwise alignments were conducted of all these TEs versus a custom BLAST database containing all *COPIA* family TEs from *A.*

thaliana and *A. lyrata*. As, this search did not led to a significant similarity to any described *COPIA* element they were considered as novel TE family which was named *HEAT ACTIVATED TRANSPOSABLE ELEMENT* (*HATE*), reflecting the fact that all 6 identified elements of this family within the *A. lyrata* genome showed transcriptional response to heat. Structural analysis suggests that all *AlHATEs* are putatively autonomous TEs. *HATE* coding sequence features the characteristic composition of *COPIA* family TEs (Figure: 3-9), where the three POL-domain proteins are arranged as INT, RT and RH (RNase H1), and are flanked by typically-oriented LTRs (Figure: 3-9). Furthermore, regulatory protein binding sequence (PBS) and polypurine tract (PPT) were identified. However, *HATE*3 has a SNP in the PPT (Figure: 3-9).

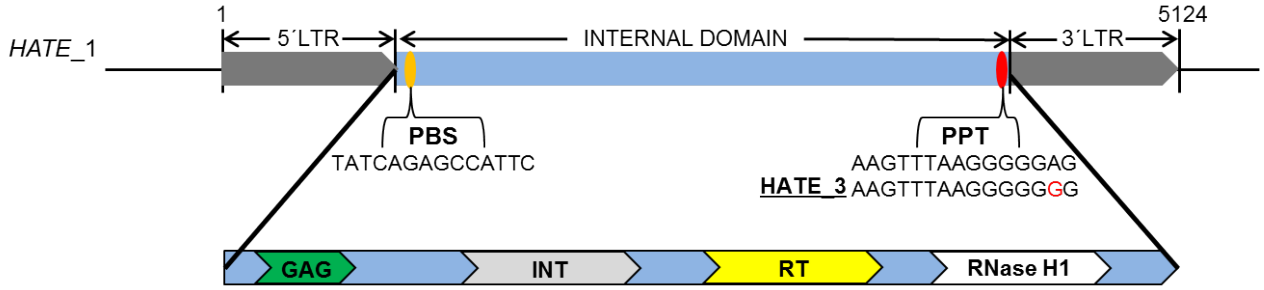

Figure 3-9: Composition and structural features of *HATE* in *A. lyrata*. The internal domain contains a Capsid-protein (GAG), integrase (INT), reverse transcriptase (RT) and RNase H1. All *AlHATE* possess a characteristic protein binding site (PBS) and a polypurine tract (PPT).

3.2.5 *ONSEN* retrotransposons in *A. lyrata*

In total, 55 *COPIA78/ONSEN* TEs were identified by automated annotation of *A. lyrata* genome (Appendix Table A5). This included 10 full-length and putatively autonomous and 45 incomplete *ONSEN* elements. A consensus sequence reconstructed from all full-length *AlONSEN* copies had a size of 5164 bp, including two flanking LTRs with each a size of 472 bp. This is longer than the consensus sequence of *AtONSEN* 4956 bp and its 440 bp LTRs. The 44 incomplete *AlONSEN* elements include solo LTRs and truncated elements missing LTRs or parts of the coding region. Within these study, all 10 full-length TEs and 8 solo LTRs were transcriptionally activated by heat treatment (Table: 3-2).

Table 3-2: Table of transcriptionally up-regulated *ONSEN* TEs in *A. lyrata* after 6 hours of heat. Transcriptional activity is represented as the log2fold-change relative to mock treated plants. Significance is stated as p-value adjusted for multiple testing with the Benjamin-Hochberg procedure. [1] full-length element.

Name	Scaffold	Start	End	Log2 fold-change	Adjusted p-value
AlONSEN1 [1]	1	11269188	11273881	7.04	4.03E-30
AlONSEN2	1	24919206	24919585	4.28	9.12E-08
AlONSEN3	2	1120762	1121196	4.94	2.76E-11
AlONSEN4 [1]	2	4268383	4273330	6.60	2.17E-25
AlONSEN5	2	12788861	12789817	5.30	4.06E-05
AlONSEN6 [1]	3	13033504	13038475	9.05	3.03E-45
AlONSEN7 [1]	3	14350695	14355704	10.50	1.85E-61
AlONSEN8	3	23055604	23059108	8.20	1.33E-34
AlONSEN9 [1]	4	16304811	16309766	8.50	1.40E-38
AlONSEN10	5	4327023	4327410	3.91	5.58E-06
AlONSEN11 [1]	5	9883484	9888427	7.25	9.35E-34
AlONSEN12 [1]	6	22653438	22658398	5.42	1.28E-11
AlONSEN13 [1]	7	23781774	23786710	7.17	2.15E-32
AlONSEN14 [1]	8	15966326	15970199	5.96	4.30E-39
AlONSEN15 [1]	247	3711	8654	7.47	5.97E-37
AlONSEN16	1007	2411	3819	7.08	2.64E-31
AlONSEN17	638	3577	4818	8.85	8.55E-54
AlONSEN18	7	8283764	8284127	3.12	1.21E-02
AlONSEN19	1007	42	1687	6.58	5.61E-25

3.2.6 Prolonged transcriptional activity of heat induced TEs in *A. lyrata* after recovery

Differential TE expression was analyzed for 6 hours heat and 48 hours recovery relative to mock. In order, to reveal TEs with prolonged heat induced activity after 48 h of recovery, overlapping TEs of both datasets were displayed by a Venn diagram. However, no significantly differentially expressed TEs were identified after recovery relative to mock in *A. thaliana*. This indicated re-establishment of silencing after

recovery (Pecinka et al., 2010). In contrast, the intersection of both datasets in *A. lyrata*, revealed five TEs that remained significantly up-regulated by heat and after 48 hours recovery (Figure: 3-10; Table: 3.3).

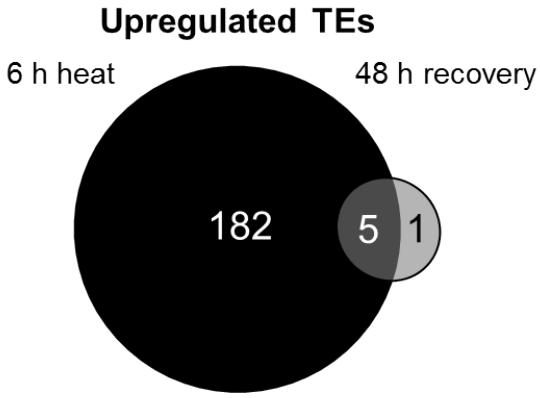

Figure 3-10: Venn diagram representing significantly up-regulated TEs after 6h at 37 °C heat (n=187) and after 48 h of recovery at ambient temperature (n=6) in *A. lyrata*. Intersection of both datasets indicates prolonged heat responsiveness.

The strongest response to heat and an almost equal transcriptional activity after recovery was observed in the two *COPIA* family TEs *AlCOPIA37* and *ATRE1* (Table: 3-3). The almost unchanged transcriptional activity of both TEs suggested good candidates for possible transposition in wild-type *A. lyrata* MN47. Analysis of *ATN9_1* revealed that this element resides within *A. lyrata OXOPHYTODIENOATE-REDUCTASE 3* (*OPR3*), an ortholog of *AtOPR3* (AT2G06050), which is involved in jasmonic acid (JA) and 12-oxophytodienoic acid (OPDA) biosynthesis (Clarke et al., 2009). This indicated that *ATN9_1* mRNA found in this study potentially represents a read through transcript from *AlOPR3* or can be activated owing to its overlap with gene position. Similar finding was made at *ALINE1_3A*, which is located within a plasma membrane-localized Remorin family protein of unknown function orthologous to AT2G45820. Thus, it can be assumed that transcriptional TE activity is potentially associated with corresponding gene expression (Table: 3-3). Non-autonomous *Sadhu6_1* (SINE) decreases after recovery but indicates still significant upregulation relative to mock (Table: 3-3), which could be due to delayed silencing of *Sadhu6_1* element. In contrast to *A. thaliana*, TEs with prolonged heat induced activity were found in *A. lyrata* post-stress recovery.

Table 3-3: Significantly up-regulated TEs after 48 hours of recovery from heat treatment. Transcriptional log2-fold change is given as treatment relative to mock.

			Log2FoldChange		
Family	**Scaffold**	**Start..end**	**6 hours heat**	**48 hours recovery**	**Notes**
AlCOPIA37 *(COPIA)*	7	7640274..7644088	7.41	6.77	Full-length element
ATRE1 *(COPIA)*	4	20662423..20666571	6.30	6.42	Full-length element
ATN9_1 *(DNA / MuDR)*	3	14608387..14608786	5.31	4.37	Within AL3G42540
Sadhu6-1 *(SINE)*	5	13548481..13549298	6.20	3.31	
LINE1_3A *(LINE)*	4	22298701..22299269	2.70	3.93	Within AL4G43220

3.3 Evolution of *ONSEN* heat responsiveness

3.3.1 Analysis of *ONSEN*-LTRs indicate common HREs and allelic variation in *A. lyrata*

Whole transcriptome analysis indicated massive heat-inducibility of *ONSEN* TEs in both, *A. thaliana* and *A. lyrata*. Recent study showed that heat responsiveness of *AtONSEN* is controlled by the *HSFA1*-pathway and mediated by DNA-binding transcription factor *HSFA2*, which recognizes heat responsive DNA elements (HREs) within the LTR (Cavrak et al., 2014). At least three repeated nGAAn boxes are needed to allow binding of a HSF trimer (Wu, 1995; Enoki et al., 2011). In *A. thaliana*, *ONSEN* harbors four nGAAn motifs that can constitute two alternative (A and B) HSF binding sites (Figure 3-11). As this was indicated as a crucial factor for transcriptional response to heat in *AtONSENs,* further comparisons focused on the evolution of HREs within the *ONSEN*-LTRs in *A. lyrata and A. thaliana.* Therefore, *ONSEN*-LTR sequences were extracted and analyzed.

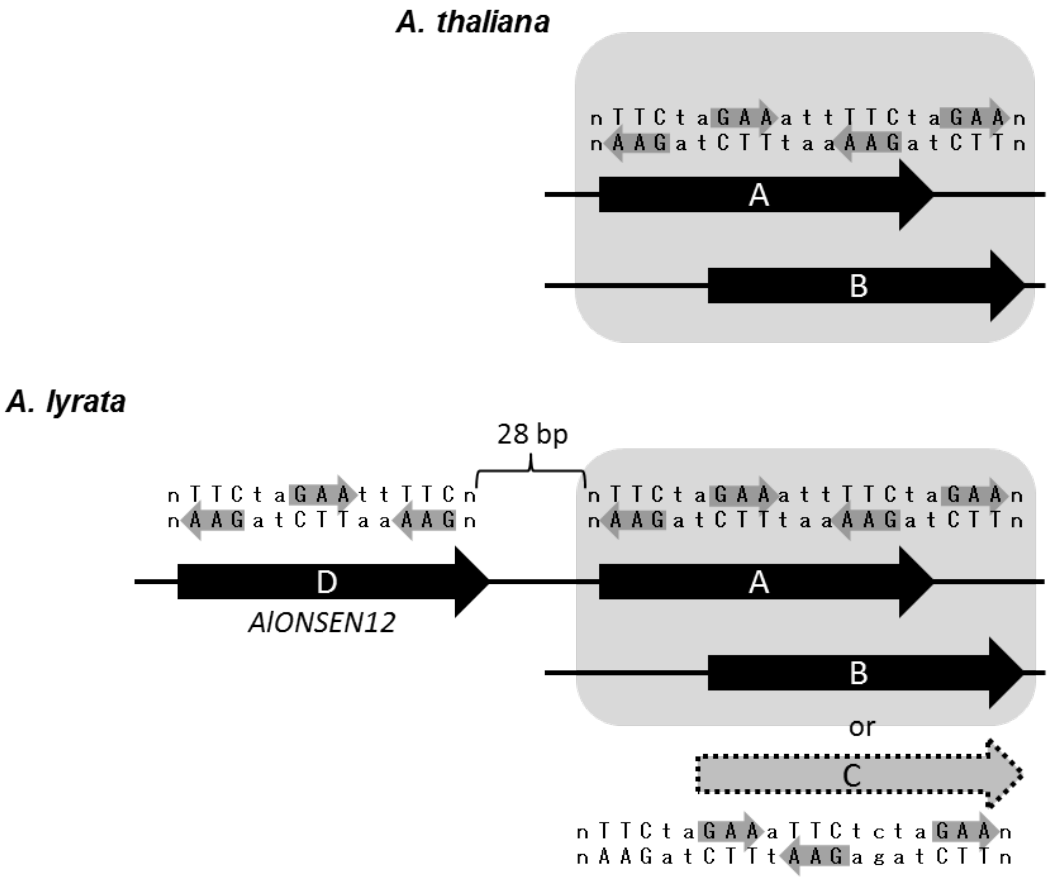

Figure 3-11: Schematic representation of HREs within *A. thaliana* and *A. lyrata ONSEN*-LTRs. The grey colored frames represents the "core HREs" identified by Cavrak et al. (2014) in *A. thaliana*. Arrows indicate HRE binding within three nGAAn boxes for high affinity HSF recognition. Insertions in *A. lyrata* C-type HRE are in red.

In seven out of eight *AtONSEN* copies, both HRE binding sites were conserved. The only exception was *AtONSEN*4, which contains a single base substitution $G_{150}C$ (Table 3-4). However, this is expected to compromise the fourth nGAAn motif and HRE B, but should not affect the high affinity HRE A (Cavrak, 2014).

Table 3-4: Heat responsive elements (HRE) composition in *A. thaliana* and *A. lyrata ONSEN* LTRs. [1] From Cavrak et al. (2014); [2] Incomplete in *A. lyrata* genome assembly.

Name	LTR length	HRE composition
A. thaliana		
AtONSEN1 [1]	440/440	A-B
AtONSEN2 [1]	440/440	A-B
AtONSEN3 [1]	463/436	A-B
AtONSEN4 [1]	436/440	A
AtONSEN5 [1]	440/440	A-B
AtONSEN6 [1]	440/440	A-B
AtONSEN7 [1]	440/437	A-B
AtONSEN8 [1]	440/441	A-B
A. lyrata		
AlONSEN1	444/444	A-C
AlONSEN2	429 soloLTR	A-C
AlONSEN3	440 soloLTR	No complete HRE
AlONSEN4	446/424	A
AlONSEN5	447	A-C
AlONSEN6	449/449	A-B
AlONSEN7	442/442	A-B
AlONSEN8	432[2]	No complete HRE
AlONSEN9	442/442	A-B
AlONSEN10	388 solo LTR	A-B
AlONSEN11	444/444	A-B
AlONSEN12	444/444	D-A-C
AlONSEN13	432/432	No complete HRE
AlONSEN14	395/442	A-B
AlONSEN15	443/443	A-B
AlONSEN16	386[2]	A-B
AlONSEN17	442[2]	No complete HRE
AONSEN18	-	-
AlONSEN19	-	-

Subsequently, *AlONSEN* LTRs were analyzed for presence of HREs. HREs were not found in two putatively non-autonomous elements *AlONSEN18* and *19* which were completely missing LTRs and parts of the coding region due to incomplete sequencing information (*AtONSEN18 and 19*). In 14 out of 19 heat responsive *AlONSEN,* at least one high affinity HRE was detected. Six full-length *AlONSEN* (*AlONSEN6, 7, 9, 11, 14, 15*) and two solo LTRs (*AlONSEN10* and *16*) contained *A.*

thaliana-like A and B-type HREs (Table: 3-4; Figure: 3-11). This indicated potentially strong evolutionary conservation of these HREs in both species. In addition, an alternative high affinity HRE (C-type) was found in two full-length elements (*AlONSEN1, 12*) and two solo-LTRs (*AlONSEN2 & 5*). The C-type largely overlaps with the B-type and differs by a single base substitution $T_{145}C$, which leads to a shift in the alignment of the nGAAn motifs (Figure: 3-12). In addition to the C-type motif, the LTR of *AlONSEN12* hold an additional putative high affinity HRE (nTTCtaGAAnnTTCn; D-type) 28 bp upstream of the "A"-type HRE motif (Figure: 3-13; Table: 3-5). In contrast, substitutions of $A_{142}G$, $T_{143}A$ and $T_{145}C$ within *AlONSEN17 and 3* led to an interruption of both HREs, missing each of one third HSF recognition site (Figure: 3-12). Besides, *AlONSEN8* and *13* lacked the whole first palindromic repeat (Table: 3-4) but, a putative third nGAAn HSF recognition site resides 5 bp more downstream (Figure: 3-12). However, these represented minor frequency alleles within *A. lyrata*.

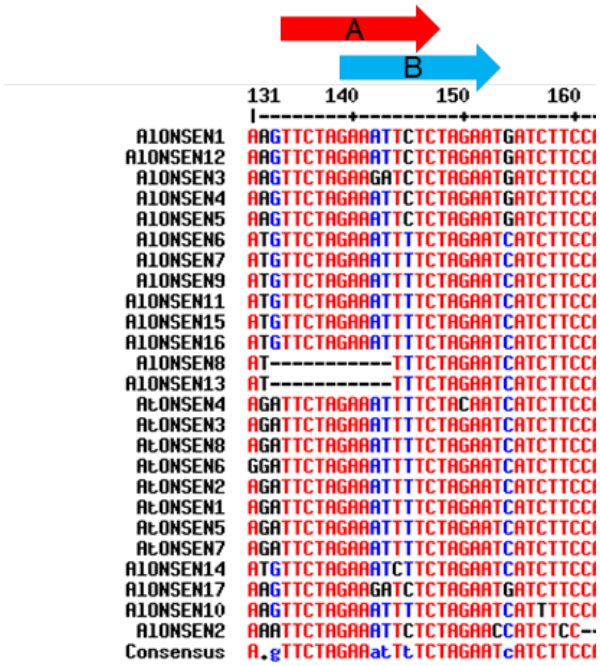

Figure 3-12: Multiple sequence alignment using hierarchical cluster analysis of the A- and B-type HRE binding sites in *A. thaliana* and *A. lyrata ONSEN*. Red letters indicate high conservation, blue letters medium conservation and black letters low conservation.

The evolutionary history between *ONSEN* from *A. thaliana* and *A. lyrata* was inferred from their 5´LTR sequences by using the Neighbor-joining method (Saitou et al., 1987). A bootstrap consensus tree inferred from 1000 bootstraps was taken to represent the evolutionary history of *A. thaliana* and *A. lyrata ONSEN* (Felsenstein,

1985). The evolutionary distances were computed using the Kimura-2 parameter method (Kimura, 1980). In addition, insertion times of *ONSEN* retrotransposons were estimated by the percentage base-pair difference of the LTRs.

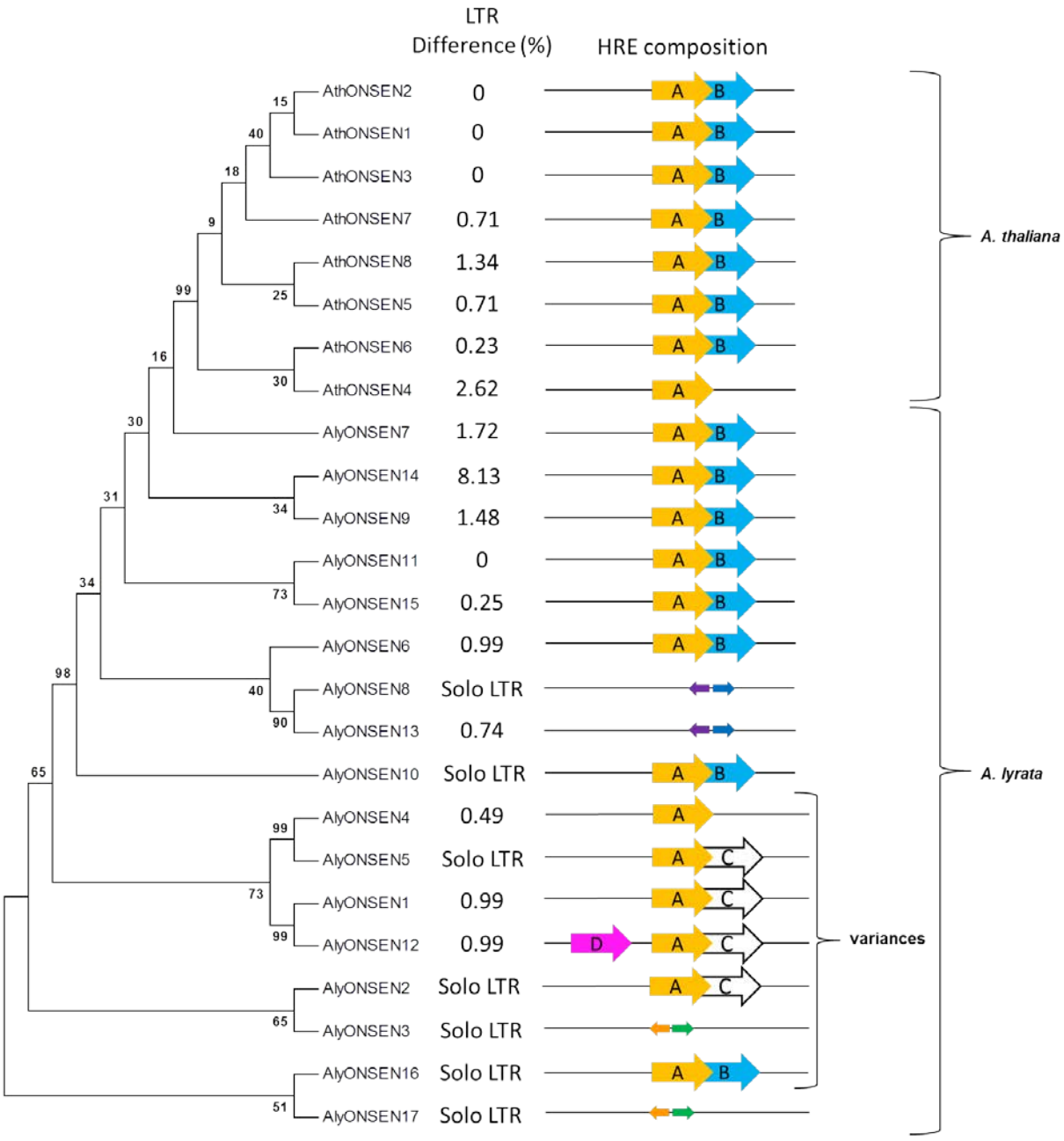

Figure 3-13: Bootstrap consensus tree from 1000 bootstrap replicates, representing the evolutionary history inferred by Neighbor-Joining method of *A. thaliana* and *A. lyrata* *ONSEN*-LTRs (Felsenstein, 1985; Saitou et al., 1987). Evolutionary distance was computed using Kimura-2 parameter method (Kimura, 1980). LTR difference was estimated by pairwise basepair (bp) difference between 5´ and 3´ LTR in autonomous *ONSEN*. HRE indicated by yellow (A) and blue (B) white (C) or pink (D) arrows. *ONSEN* without HRE, but conserved motifs are indicated in small colored arrows reflecting first (orange), second (green), third (violet) and fourth (darkblue) HSF recognition site (nGAAn-box).

A. thaliana and *A. lyrata ONSENs* formed separate clades based on their LTR sequences (Figure 3-13). Within the *A. thaliana* clade, comparison of *AtONSEN1*, *AtONSEN2* and *AtONSEN3* indicated a 100% sequence similarity of 5' and 3' LTR sequences which pointed out recent origin of these copies (Figure: 3-13). The small bootstrap probabilities of the external branches diverging *AtONSEN1, 2, 3, 5, 7,* and *8* demonstrated a less strict divergence due to high sequence similarities between the *AtONSEN*. However, *AtONSEN6* and *4* showed a strict divergence from the other *AtONSEN* with a 99 % bootstrap probability (Figure 3-13). As *AtONSEN6* indicated being younger than *AtONSEN8* (0.23 versus 1.43 % bp difference; Figure: 3-13) but strictly related to *AtONSEN4*, this led to the assumption, that *AtONSEN6* is potentially a transposed copy of *AtONSEN4*. The presence of the $C_{152}G$ SNP in *AtONSEN*4, leading to the loss of B-type motif, but not in *AtONSEN6* suggested that this evolved potentially after the transposition.

The *A. lyrata* clade demonstrated a difference of the *AlONSEN* containing A- and B-type HREs and variants like C-type or incomplete HREs i.e. single HSF recognition sites, by a bootstrap probability of 98% (Figure 3-13). *AlONSEN*-LTRs with the A- and B-type HRE motif were relatively older than those including variant type motifs. Indicated by the low LTR divergence, recent transpositions of *AlONSEN* containing the A- and C-variant (*AlONSEN 1, 12*) were suggested.

In summary, the analysis of *A. lyrata ONSEN* LTRs revealed two major, partially overlapping HREs, representing high affinity HSF binding sites. The A-type motif was more conserved in *A. lyrata* and was present in all *ONSENs* of *A. thaliana*. Hence, LTR analysis indicated presence of evolutionary conserved HREs in the common progenitor of *A. thaliana* and *A. lyrata*. The phylogenetic reconstruction of *ONSEN* in *A. thaliana* and *A. lyrata* led to the suggestion that HREs with A- & B-type HSF binding site is the ancestral due to (*i*) presence in both species and (*ii*) the younger TE age of *ONSEN*-LTRs consisting HRE variants.

3.3.2 *ONSEN* elements are present in other *Brassicaceae* species

The reconstruction of the evolutionary history of *ONSEN*-LTRs in *A. thaliana* and *A. lyrata* raised new questions about the development of the C-type motif as well as the evolution of the HRE consisting of A- and B-type HREs. Further analyses were conducted by investigating *ONSEN* in other *Brassicaceae* species for the presence of HRE and reconstructing its evolutionary conservation.

COPIA78/*ONSEN* elements were extracted by genome-wide blast search in Phytozome v10 (Goodstein et al., 2012) from publicly available *Brassicaceae* genomes of *Boechera stricta* v1.2 (DOE-JGI, http::/www.phytozome.net/bstricta), *Brassica rapa* FPsc v1.3 (DOE-JGI, http::/www.phytozome.net/BrapaFPsc), *Capsella grandiflora, C. rubella,* (Slotte et al., 2013)*, Eutrema salsugineum* (Yang et al., 2013). In addition, custom BLAST databases were created of *Arabis alpina* (Willing et al., 2015) and *Ballantinia antipoda* (Vu, Finke, Pecinka; unpublished data) genomes and observed for *ONSEN* homologs using BLAST search tool within BioEdit (Hall, 1999) and CLC workbench (http://www.clcbio.com)**.** Subsequently, LTRs were reconstructed and analyzed for the presence of HREs. In agreement with previous study (Ito et al. (2013), no *ONSEN* elements were detected in *C. grandiflora* and *C. rubella*. However, putative *ONSEN* elements were identified in other species (Table: 3-5)*.*

Table 3-5: List of putative *ONSENs* in *Arabis alpina*, *Ballantinia antipoda*, *Boechera stricta*, *Brassica rapa*, *Eutrema salsugineum*. HRE were estimated by sequence identities to known *ONSEN* HRE.

Name	Location Chromosome:start..end	LTR size	nGAAn clusters
Arabis alpina			
AaONSEN1	scaff_50914_1:21202 .. 21347	231	nTTCtaGAAn[5bp]nGAAn
Ballantinia antipoda			
BaONSEN1	Contig_1234567: 212152	341	nTTCtaGAAnnnTTCtaGAAn
BaONSEN2	Contig_1234567: 122839	326	nTTCtaGAAn
BaONSEN3	Contig_1567: 193145	137	nTTCtaGAAnnnTTCtaGAAn
Boechera stricta			
BsONSEN1	scaffold_24340: 30478 .. 30921	443	nTTCtaGAAn
BsONSEN2	scaffold_24340: 86763 .. 87180	417	nTTCtaGAAn
Brassica rapa			
BrONSEN1	A09: 1408383972 .. 14087544	387	nTTCtaGAAn
BrONSEN2	A04: 600303 .. 605484	451	-
BrONSEN3	A08: 8071678 .. 8076523	-	nTTCtaGAAn[3bp]nTTC
BrONSEN4	A10: 12639426 .. 12642605	-	-
BrONSEN5	A01: 14603742 .. 14603324	-	-
BrONSEN6	A05: 22005584 .. 22009871	-	-
Eutrema salsugineum			
EsONSEN1	Scaffold_14: 2891845 .. 2892295	450	nTTCnnGAAn[42bp] nTTCnnGAAn
EsONSEN2	Scaffold_3´: 6431065 .. 6433482	-	-

HSF recognition sites were identified in 2 out of 6 *B. rapa ONSEN* (*BrONSEN*). However, high affinity binding can only be established in *BrONSEN3* due to an additional nGAAn motif 3 bp downstream of the palindromic repeat (Wu, 1995).

In *A. alpina (AaONSEN)*, only a single, putative *ONSEN* solo LTR was identified. This *AaONSEN* contained a single palindromic repeat (nTTCtaGAAn) with a putative third nGAAn-box 5 bp downstream, indicating presence of a gap-type HRE ((Sakurai et al., 2010);Table: 3-5). In contrast, two incompletely assembled, but most likely full length, *ONSENs* and one solo LTR were identified in *B. antipoda*. The LTR in one autonomous *B. antipoda ONSEN* (*BaONSEN1*) and in the solo LTR (*BaONSEN3*) perfectly recapitulated the A- and B-type HRE of *ONSEN* in *A. thaliana* and *A. lyrata*. However, in *BaONSEN3* one palindromic repeat was absent (Table: 3-5). Two autonomous *ONSEN* retrotransposons were discovered in *B. stricta* (*BsONSEN*), located at one scaffold. Both identified *BsONSEN* LTRs consists of only one palindromic (nTTCtaGAAn) repeat (Table: 3-5). In *E. salsugineum*, two *ONSENs* were found, but only one featured two times a palindromic nTTCtaGAAn motif within

its LTR. However, these are separated by a 42 bp gap which excluded HSF recognition.

In summary, outside of genus *Arabidopsis* we found HREs only in two copies of *BaONSEN*. In all other observed species, *ONSEN*-LTRs included only single palindromic repeat (nTTCtaGAAn). This led to the assumption that the complex A- and B-type HRE evolved out of a single palindromic repeat, in which HSF recognition and binding was inefficient (Wu, 1995; Sakurai et al., 2010).

3.3.3 Phylogenetic shadowing indicates evolution of *ONSEN* HREs from a single palindromic sequence

In order to reconstruct the evolution of cis-regulatory HREs in *ONSEN*, conservation of the LTR sequences among investigated species were calculated by phylogenetic shadowing. Because of the different *ONSEN* abundance within the *Brassicaceae* species, which could lead to a stronger emphasis of species with high *ONSEN* copy number, species-specific consensus sequences were generated. AVID alignments and computation of sequence similarity, was carried out against the consensus LTR sequence of *AlONSEN* (Bray et al., 2003).

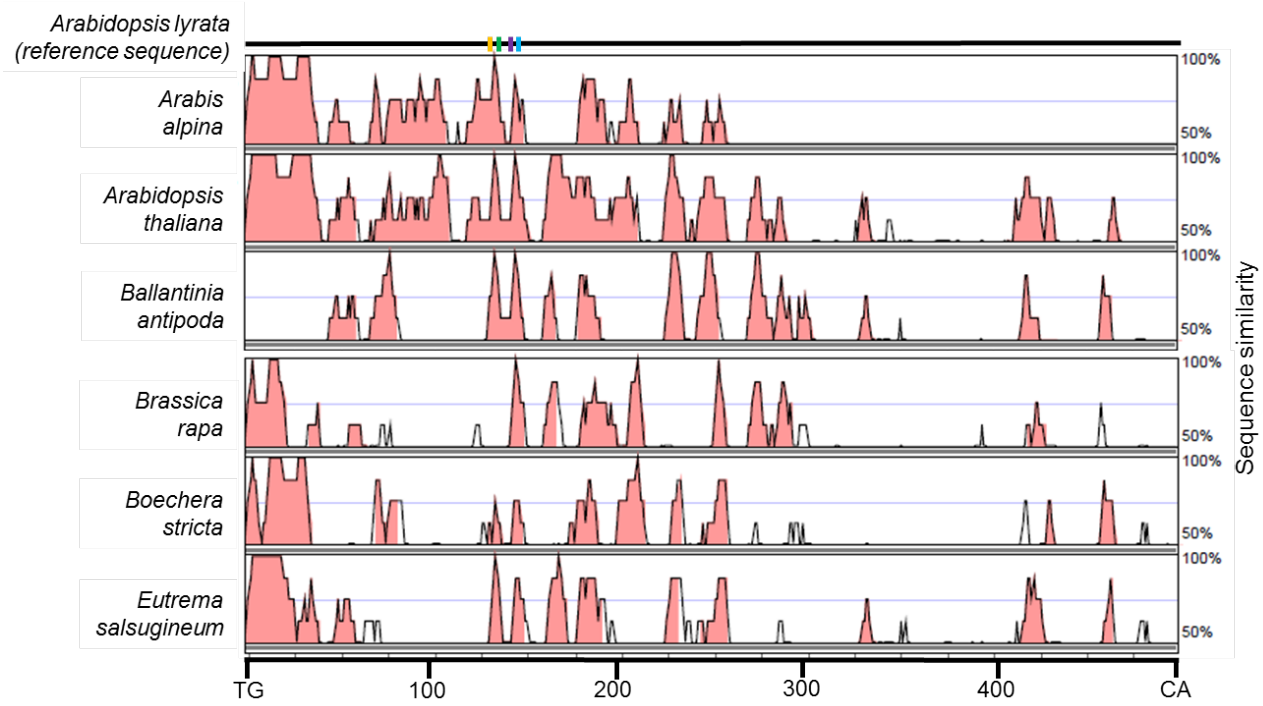

Figure 3-14: Phylogenetic shadowing of *ONSEN*-LTRs of *ONSEN* homologs, isolated from seven *Brassicaceae* species. The *ONSEN*-LTR of *Arabis alpina*, *Arabidopsis thaliana*, *Ballantinia antipoda, Brassica rapa* and *Boechera stricta* were each aligned to the *ONSEN*-LTR of *A. lyrata*. All pairwise alignments are presented as VISTA plots. Pink color indicates regions with at least 70% of conservation on an 8-bp sliding window. HSF recognition sites (nGAAn) are marked in reference dash by vertical dashes, reflecting first (orange), second (green), third (violet) and fourth (darkblue) HSF recognition site (nGAAn-box).

Phylogenetic shadowing indicated a 100% conservation rate of all four HSF recognition sites in *A. thaliana* and *B. antipoda* (Figure: 3-14). In contrast, the *ONSEN*-LTR sequences of *A. alpina* and *E. salsugineum* indicated 100% sequence identity at the first two HSF recognition sites (Figure: 3-14). In addition, the *ONSEN*-LTR sequence of *A. alpina* showed a ~ 70 % conservation at the position of the third and fourth HSF recognition site. This conservation was favored by the putative HSF binding site (nGAAn) 5 bp downstream from the first two (3.3.2; Table: 3-5). In contrast, LTR consensus sequence of *E. salsugineum* indicated two palindromic repeats (nGAAtaTTCn) with a 42 bp gap in between. The AVID alignment aligned one palindromic repeat 135 bp downstream from TE start site with 100 % conservation to the first two HSF recognition sites. The other palindrome, found 42 bp upstream of the conserved one, indicated no sequence conservation at this position. A ~ 70 % sequence conservation at the position of the other two HSF recognition sites in *E. salsugineum ONSEN*-LTR, was indicated by phylogenetic

shadowing. However, this conservation is associated to base pairs within the calculation window and does not reflect putative HSF binding motifs.

In *B. stricta,* phylogenetic shadowing showed partial sequence conservation (~ 70 %) at all four HSF recognition sites (Figure: 3-14). The conservation sliding window barely reached the 70 % threshold. However, the palindromic repeat (nTTCtaGAAn) found in *BsONSEN* was aligned to the first two HSF recognition sites of *AlONSEN*, but the variance of the adjacent sequence led to the calculation of a lower conservation window (Appendix Figure A1). Strong sequence conservation was indicated at the third and fourth HSF binding site in *B. rapa ONSEN*-LTR (Figure: 3-14). However, a close look at the AVID alignment (Bray et al., 2003) showed that this is due to the settings of the alignment algorithm (Figure: 3-15).

Figure 3-15: Detail of the AVID alignment of *ONSEN*-LTRs of *A. lyrata* (reference) versus *B. rapa*. HSF recognition sites are marked by colored arrows, reflecting first (orange), second (green), third (violet) and fourth (darkblue) HSF recognition site (nGAAn-box). Background colors per base by default: Thymin (T; lightgreen); Adenin (A; blue); Guanin (G; violet), Cytosin (C; darkgreen), gaps and any Nukleotide (N; grey).

In summary, sequence conservation was revealed by phylogenetic shadowing of the first and second HSF recognition site in *A. thaliana*, *B. antipoda* and *A. alpina,* which featured at least three HSF recognition sites within their LTR. However, high affinity HREs were conserved in *A. thaliana*, *A.lyrata* and *B. antipoda*. In contrast, species in which only a single palindromic repeat was identified, conservation was lower (*B. stricta*) or promoted by adjacent sequences (*B. rapa*).

Here, the conservation was incorporated into a diagramed phylogenetic tree made of established *Brassicaceae* phylogeny (Mandáková et al., 2010; Franzke et al., 2011; Heenan et al., 2012). This suggests that the HREs within *ONSEN*-LTRs, evolved from a single palindromic repeat, which contains two nGAAn-motifs, via local duplication to that HRE with two HSF recognition sites, constituted by four nGAAn motifs, before the split of the genera *Arabidopsis* and *Ballantinia (*Figure: 3-16).

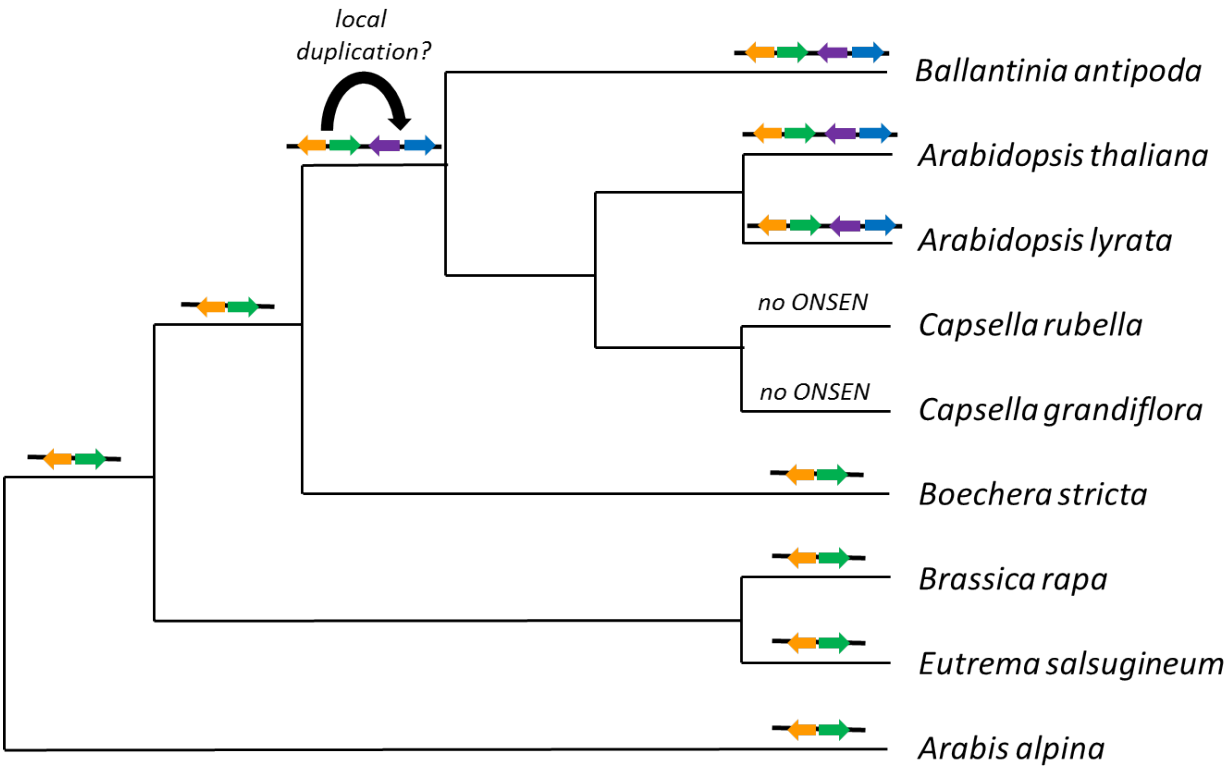

Figure 3-16: Diagramed phylogenetic tree of proposed *ONSEN* HRE evolution within observed *Brassicaceae* species. Arrows mark the palindromic repeats. It is proposed that a single palindromic repeat was present in early ancestral *ONSEN* LTR that underwent a local duplication leading to perfect high-affinity HRE in the genera *Arabidopsis* and *Ballantinia*.

3.4 Characterization of novel heat-responsive *COPIA* families

Whole transcriptome analysis revealed novel heat responsive TE families *COPIA20* and *HATE* in *A. lyrata* as well as *COPIA37* in *A. lyrata* and *A. thaliana* (Table: 3-6). Here, possible molecular basis of heat responsiveness within the *COPIA* superfamily were analyzed.

Table 3-6: List of novel heat responsive TEs, identified by whole transcriptome analysis in *A. lyrata* and *A. thaliana*. Transcriptional log2fold-change is given as treatment relative to mock. Significance is stated as p-value adjusted for multiple testing with the Benjamin-Hochberg procedure.

Family	Scaffold / chromosome	start	End	log2Fold-change	p-value
A. lyrata					
COPIA20					
AlCOPIA20_1	10	141350	143663	2.83	3.23E-03
AlCOPIA20_2	3	24168379	24172931	2.60	1.20E-02
AlCOPIA20_3	10	149249	153800	3.89	4.52E-06
AlCOPIA20_4	10	159963	161320	3.18	2.41E-02
AlCOPIA20_5	10	294152	299187	3.50	5.72E-04
AlCOPIA20_6	10	455857	459788	3.48	1.39E-04
AlCOPIA20_7	11	214101	218623	4.58	3.85E-08
AlCOPIA20_8	10	3525	8063	2.54	1.66E-02
AlCOPIA20_9	28	61453	66005	2.14	3.54E-02
COPIA37					
AlCOPIA37_1	3	24128553	24130918	3.32	2.20E-04
AlCOPIA37_2	7	7640274	7644088	7.41	5.06E-24
AlCOPIA37_4	8	22069407	22070098	3.51	1.82E-09
AlCOPIA37_5	8	22068138	22068966	2.68	7.62E-07
AlCOPIA37_6	7	20963365	20968153	2.60	2.24E-03
AlCOPIA37_7	7	24639790	24641218	1.25	2.31E-03
HATE					
AlHATE1	1	25447360	25449935	8.22	1.48E-45
AlHATE2	4	15135433	15137933	6.27	3.73E-22
AlHATE3	5	9945065	9947640	8.01	4.76E-43
AlHATE4	7	8285793	8288369	7.85	7.22E-41
AlHATE5	8	6161052	6163595	4.98	3.40E-11
AlHATE6	8	15680837	15685660	7.26	1.99E-33
A. thaliana					
COPIA37					
AtCOPIA37_1	4	1875175	1875724	2.91	1.88E-10
AtCOPIA37_2	3	10396358	10396888	2.74	1.86E-04
AtCOPIA37_3	1	3895125	3895672	2.22	3.40E-04
AtCOPIA37_4	3	10400031	10400574	2.18	2.03E-04

Sequence analysis of heat responsive *AlCOPIA20* family TEs revealed a putative HRE within their 5´LTR. The identified HRE resides approximately 86 bp from the LTR start position and has a similar sequence composition like the “B”-type HRE of *ONSEN* (Figure: 3-17; 3.3.1)

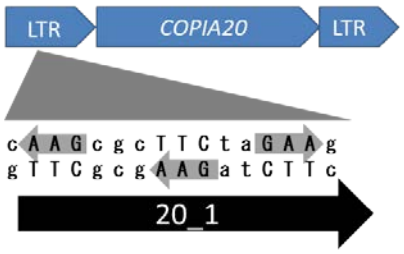

Figure 3-17: Representation of *A.lyrata COPIA*20 HRE composition within its LTRs. Grey arrows mark nGAAn HSF recognition sites. Black arrow represents putative HRE.

Whole transcriptome sequencing analysis revealed common heat responsiveness of *COPIA37* in *A. thaliana* and *A. lyrata*. In addition, *AlCOPIA*37_2 showed prolonged heat activity in *A. lyrata*. Sequence analysis of the LTRs of heat responsive *AlCOPIA37_2* revealed two putative HREs. The putative *COPIA37* HREs resides between 164 bp and 205 bp upstream from transcription start site (TSS) (Figure: 3-18; 37_1). The second motif was identified 11 bp downstream from the first putative HRE (Figure: 3-18; 37_2). However, the second motif was absent in *AlCOPIA37_1*. In addition, HREs were only identified in these two out seven heat responsive *AlCOPIA37*.

HRE motif 37_1 is composed like a step-type (nGAAn[2bp]nTTCn[2bp]nGAAn) HRE consisting of 2-bp gaps between the regulatory binding motifs. This was described in mediating basal or low level transcriptional activation (Sakurai et al., 2010; Enoki et al., 2011). HRE motif 37_2 is a gap-type (nTTCn[2bp]nGAAnnTTCn) HRE which was reported to mediate moderate transcriptional activity (Sakurai et al., 2010; Enoki et al., 2011). This indicated presence of HREs within LTRs of two *AlCOPIA37*. However, a correlation of the putative HREs to the transcriptional activation remains unclear because they are underrepresented within heat responsive *AlCOPIA37* copies. In addition, a correlation of the prolonged heat activity to the gap-type HRE 37_2 can be suggested, but needs to be further tested.

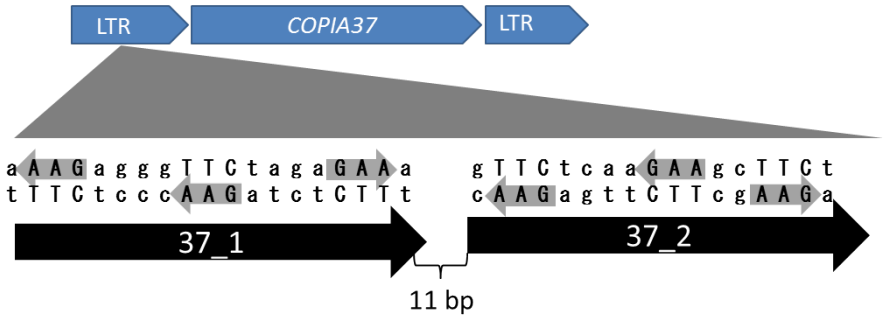

Figure 3-18: Representation of *A.lyrata COPIA37* HRE composition within its LTR. Grey arrows mark nGAAn HSF recognition sites. Black arrows represent identified step type HREs (Sakurai et al., 2010).

Sequence analysis of novel identified *HATE* family TEs for presence of HRE motifs identified two putative HRE elements within their LTRs (Figure: 3-19). However, both putative elements were separated by approximately 182 to 177 bp. In addition, both HREs differed in structure and size. The first putative HRE resided 138 bp from the transposon start site and is a variant gap-type HRE (nGAAn[2bp]nGAATTCn], consisting of three putative HSF recognition sites (nGAAn). However the two downstream HSF-binding motifs were not divided by a gap (Figure: 3-19; H1). Therefore, HSF recognition cannot be verified by *in silico* analysis. In addition, this motif is only present in 3 out of 6 *AlHATE*, it appears to be interrupted by a transition ($C_{151}T$) in last HSF-binding motif (TTCt) in *AlHATE4* and *6*.

The second motif is composed like a gap-type HRE motif (nGAAn[4bp]nGAAnnTTCn) (Figure: 3-19; H2) which were reported to mediate moderate transcriptional activity (Sakurai et al., 2010; Enoki et al., 2011). This putative HRE was observed in all *AlHATE* TEs. Thereby, composition and conservation of putative HRE H2 indicated potential HSF mediated expression of *HATE* TEs. However, HSF binding and recognition is currently unclear for putative HRE H1 and needs to be further investigated.

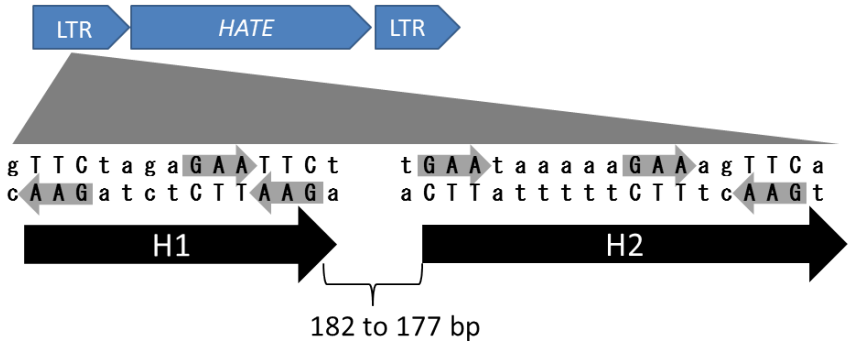

Figure 3-19: Representation of HRE composition in *A.lyrata* LTR. Grey arrows mark nGAAn HSF recognition site. Black arrows represent identified putative HREs.

In summary, analysis of the LTR sequences, of *A. lyrata* heat responsive TE families, revealed family specific putative HREs. Here, a putative *ONSEN*-like HRE was observed in the LTR of *AlCOPIA20*. In contrast, two putative HREs were identified within *AlCOPIA37*-LTRs. While HRE 37_1 was only present in two out of seven heat responsive *AlCOPIA37*, HRE 37_2 was only present in prolonged heat activated *AlCOPIA37_2*. The LTRs of the TEs from the new identified *COPIA* family *HATE* featured two putative gap-type HREs. However, HSF recognition of putative HRE H1 identified in *AlHATE* TEs, remains currently unclear, but HRE-mediated expression seems plausible for the gap-type HRE H2.

3.5 Repeated evolution of heat responsiveness

The presence of several families of heat responsive *COPIA* TEs in *A. lyrata* raised the question whether the HREs may originate from a single or multiple evolutionary events. To test this, a LTR based phylogenetic network of *A. lyrata* heat responsive and non-heat responsive TEs was constructed. By using this, evolutionary divergence of incompatible splits e.g. by recombination, duplication or horizontal transferred nuclear DNA, can be visualized by their inferred evolutionary distance (Huson et al., 2006). The incompatible splits would appear high distant with low probabilities in phylogenetic trees. Heat responsive *Gypsy* family *ATGP9B*-LTRs were used as outgroup.

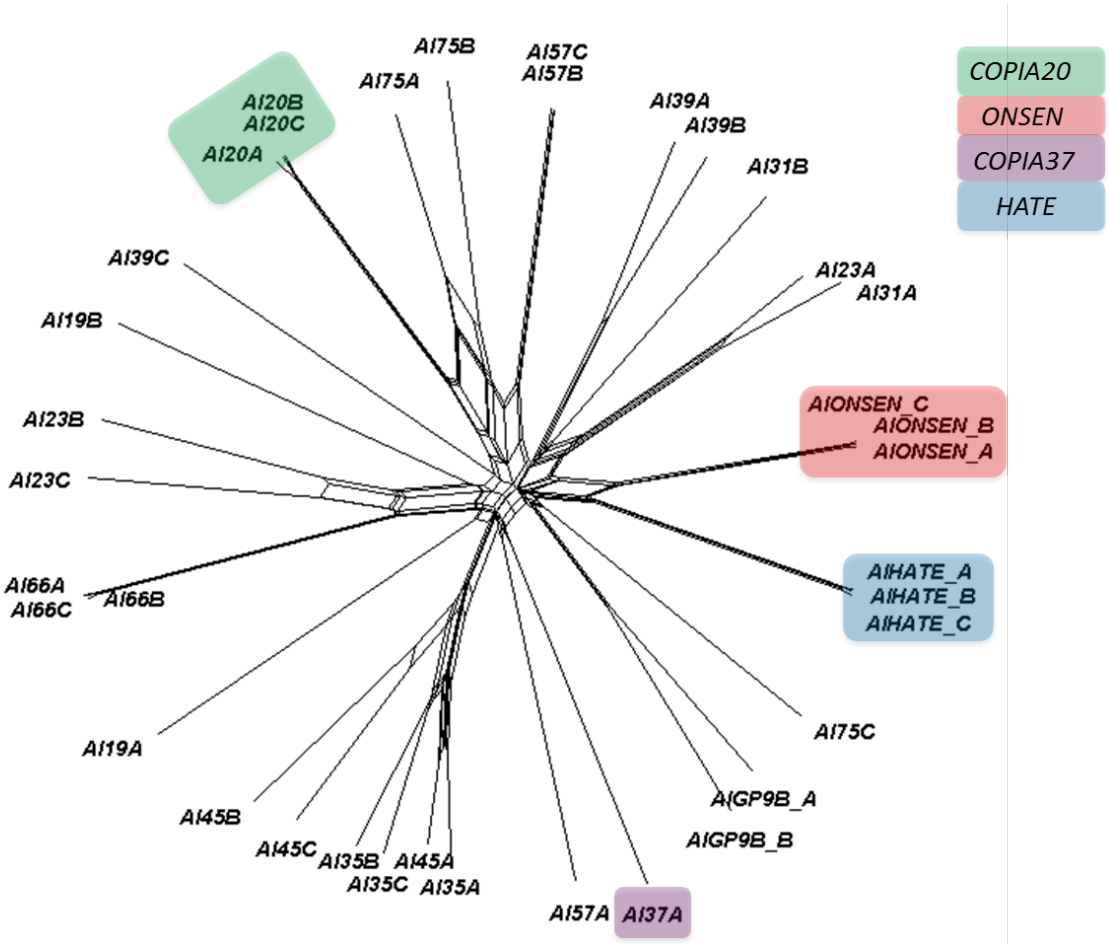

Figure 3-20: LTR based phylogenetic network representing the evolutionary distance of heat responsive and non-heat responsive *AlCOPIA* families. *Gypsy* family elements of heat responsive *AlGP9B* were used as outgroup. *AlCOPIA* heat responsive families and corresponding legend are assigned by colors.

The inferred phylogeny by a network showed a compatible split of *ONSEN* and *HATE* LTRs (Figure: 3-20). This initially indicated a potential simultaneous evolution of HREs in both TEs. However, the early split and the length of the branches, which display additional evolutionary distance, suggest that the HRE variants between these two LTRs might have evolved independently.

In contrast, *AlCOPIA20* and *AlCOPIA37* split very early from a common ancestor with *AlONSEN* and *AlHATE*. Besides, position of *AlCOPIA20* indicated high evolutionary distance from all other heat responsive TEs with much splits. This led to the assumption of an independent evolution of heat responsiveness in *AlCOPIA20*. Similar observation was done in heat responsive *AlCOPIA37*, which indicated an early split from the other heat responsive TEs. A compatible split from *AlCOPIA57* indicated common evolution of both TE families. However, *AlCOPIA57* were not observed as being heat responsive within this study

3.6 Putative novel HREs are partially conserved within homologous TEs in other *Brassicaceae* species

The phylogenetic network demonstrated a potential repeated evolution of heat responsiveness within *COPIA* families in *A. lyrata*. However, this raised the question if that observed putative HRE within *AlCOPIA*20, *AlCOPIA*37 and *HATE* are evolutionary conserved in other *Brassicaceae*. Therefore, BLAST searches were conducted on publicly available *Brassicaceae* genomes as previously described in *ONSEN* (3.3.2). Subsequently, sequences of identified *COPIA20*, *COPIA37* and *HATE* homologs were extracted and their LTR reconstructed. Afterwards, the reconstructed LTRs were observed for the presence of the corresponding HREs.

Table 3-7: List of *COPIA20*, *COPIA37* and *HATE* TEs observed in *Brassicaceae* featuring corresponding HRE in their LTR. LTR identity was observed in full-length (Full) TEs by a pairwise alignment of 5´ versus 3´LTR. HREs were assigned by sequence motif identity to *A. lyrata* homolog. Corresponding HREs were assigned by their naming in *A. lyrata* homologs, additional novel identified HREs are assigned as "X".

Name	Chr./ Scf.	Start	End	Type	LTR size (bp)	LTR identity (%)	HRE
COPIA20							
A. lyrata							
AlCOPIA20_1	10	148631	154206	full	406	99	20_1
AlCOPIA20_2	10	159541	159962	solo	422	-	20_1
AlCOPIA20_3	10	294152	300537	full	422	98	20_1
AlCOPIA20_4	11	213691	214100	solo	410	-	20_1
AlCOPIA20_5	28	61046	66626	full	407	100	20_1
AlCOPIA20_6	13	234202	239757	full	419	100	20_1
COPIA37							
A. lyrata							
AlCOPIA37-1	7	7640274	7644088	full	429	98	37_1
AlCOPIA37-2	3	24130517	24130953	solo	437	-	37_1–37_2
Arabidopsis thaliana							
AtCOPIA37_2	3	12473548	12473995	solo	448	-	37_1
AtCOPIA37_4	3	12513617	12514054	solo	439	-	37_1
AtCOPIA37_6	3	11393988	11394417	solo	432	-	37_1
AtCOPIA37_7	3	20505028	20505461	solo	435	-	37_1
Capsella rubella							
CrCOPIA37_1	1	17530487	17530921	full	409	-	37_1
CrCOPIA37_2	1	14555691	14556103	solo	409	-	37_1

Name	Chr./ Scf.	Start	End	Type	LTR size (bp)	LTR identity (%)	HRE
Capsella grandiflora							
CgCOPIA37_1	9816	17768	18180	solo	409	-	37_1
Eutrema salsugineum							
EsCOPIA37_1	1	11995597	11996037	solo	441	-	37_1 variant
EsCOPIA37_2	15	759894	7570335	solo	442	-	37_1 variant
HATE							
A. lyrata							
AlHATE_1	1	25445324	25445865	full	525	99%	H1-H2
AlHATE_2	4	15133415	15138536	full	524	100%	H1-H2
AlHATE_3	5	9943039	9948175	full	532	99%	H1-H2
AlHATE_4	7	8283767	8288904	full	532	99%	H2
AlHATE_5	8	6159034	6164131	full	524	97%	H1-H2
AlHATE_6	8	15680837	15685660	full	540	99%	H2
Boechera stricta							
BsHATE_1	26833	1058555	1059089	solo	525	-	H1
BsHATE_3	4232	581	1095	solo	515	-	X
BsHATE_4	30057	1058559	1062295	full	526	94	X – H1
BsHATE_5	10199	144407	149499	full	513	96	X – H1
BsHATE_6	13129	2245568	2246076	solo	509	-	X – H1
BsHATE_7	3288	24153	29242	full	509	96	X – H1
BsHATE_9	3148	824911	830019	full	513	97	X – H1
BsHATE_10	18473	1614326	1619381	full	513	96	X – H1
BsHATE_11	8819	608944	613973	full	494	98	variant X – H1
BsHATE_12	7867	789356	789865	solo	510	-	X
BSHATE_13	556	4969780	4974903	full	508	95	X
BsHATE_14	26959	2009750	2010261	solo	512	-	X – H1

Here, no *COPIA20* were identified in any other *Brassicaceae* species except *A. lyrata* and *A. thaliana*. Although, six *COPIA20* family TEs are annotated in *A. thaliana* neither HREs nor heat responsive *AtCOPIA20* TEs were identified within this study, because *AtCOPIA20* TEs are all truncated and lack of LTR sequence information. This loss of sequence information, truncation and low copy number of *AtCOPIA*20 led to the suggestion that they got degenerated within the genome by illegitimate recombination and unequal intra-strand homologous recombination (Devos et al., 2002; Pereira, 2004; Tenaillon et al., 2010; Sabot et al., 2011). In contrast, *COPIA20* TEs are highly abundant in *A. lyrata* with 269 copies genome-wide. The putative HRE were identified in six out of 9 heat responsive copies.

Homologs of *AlCOPIA37* and *AlHATE* were identified (Appendix Table A6). *COPIA37* TEs were identified in *A. thaliana* (n = 32), *B. rapa* (n = 2), *C. rubella* (n = 2), *C. grandiflora* (n = 1) and *E. salsugineum* (n = 2). No HREs were identified in *B. rapa COPIA37* LTR. While *A. thaliana* indicated the highest number of *COPIA37* copies within the genome, only four were transcriptionally induced to heat. However, no HREs were identified in these transcriptionally upregulated TEs. In contrast, four *COPIA37* solo LTRs that showed no upregulation to heat in performed whole transcriptome analysis featured the HRE 37_1 motif of *AlCOPIA37*. This motif was also identified within *COPIA37* in the *Capsella* genus, although *COPIA37* copies appeared in a low abundance. Besides, *COPIA37* in *E. salsugineum* features a HRE C37_1 variant which misses the upstream HSF recognition site, but features a nGAAn-box downstream of the core motif (Figure: 3-21). This nGAAn-box was identified also in *CgCOPIA37*, *CrCOPIA37* and in *AlCOPIA37_1* (Figure: 3-21). However, these indicates a variant step-type HRE and appears in other species HSF recognition cannot be excluded.

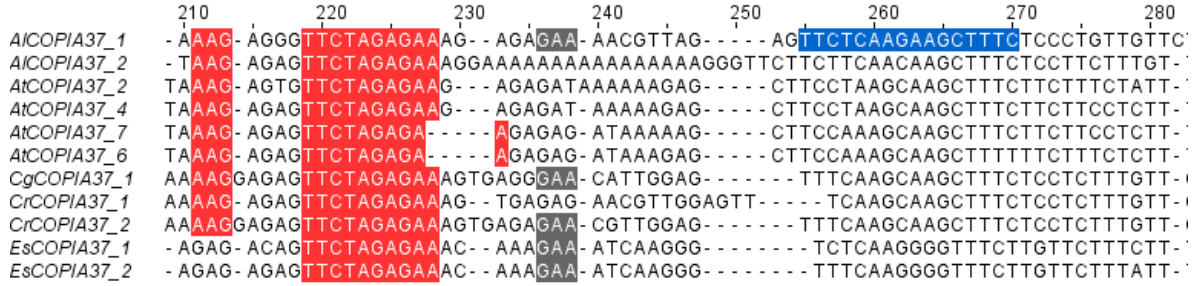

Figure 3-21: Detail of the HRE region from *COPIA37*-LTR alignment of *A. lyrata*, *A. thaliana*, *C. rubella*, *C. grandiflora* and *E. salsugineum COPIA37* homologs. Identified HREs are red (37_1) and blue (37_2) colored. Additional putative HSF recognition sites of C37_1 are colored grey.

Homologs of novel identified TE family *HATE* were identified in *A. lyrata* (n = 6), *B. stricta* (n = 14), *B. rapa* (n = 2) *and E. salsugineum* (n = 6). However, HREs were exclusively identified in *A. lyrata* and *B. stricta HATE* (*BsHATE*). Thereby, a large accumulation of *HATE* was observed in *B. stricta*, which resides in total 14 *BsHATE* TEs, while in contrast *A. lyrata* contained six *AlHATE* TEs. In 12 out of this 14 *BsHATE* candidates HREs were identified. Putative HRE H1 (Figure: 3-18; Table: 3-7) was present in 9 *BsHATE*, while the second identified HRE of *AlHATE* (Figure: 3-19; H2) was fully absent. In addition, a novel HRE that was identified in 10 *BsHATE* approximately 40 bp upstream of the HRE H1 (Table: 3-7 HRE “X”). This novel HRE

is a gap-type HRE (nTTCnnGAAn[5bp]nGAAnnTTCn) with an additional putative HSF recognition sites in between the 5 bp gap (Figure 3-22; marked violet). This putative novel HRE in *BsHATE* had two putative HSF binding sites like *ONSEN*. However, here the third HSF recognition site (nGAAn) box is always separated by a 5 bp gap, indicating two step-type HSF binding sites. These step-type HREs were postulated to mediate moderate transcriptional activation (Sakurai et al., 2010; Enoki et al., 2011).

This variance in presence of HREs within one TE family in different species emphasized the assumption that heat response evolved repeatedly and potentially independent within species.

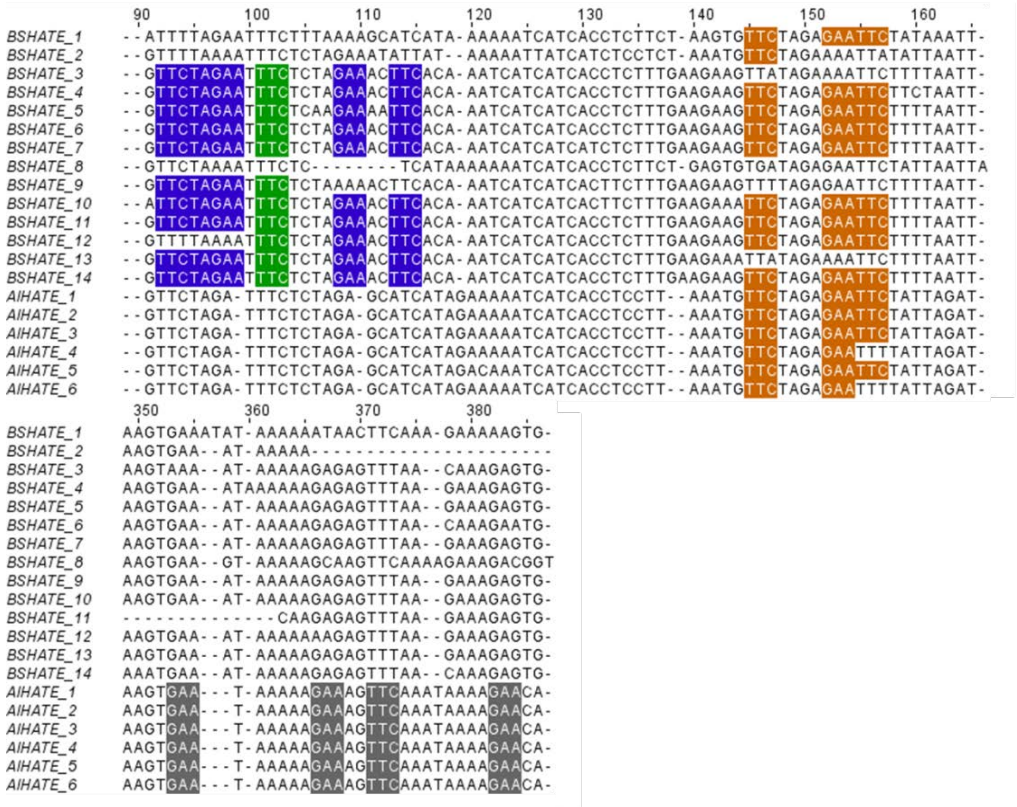

Figure 3-22: Detail of the HRE regions from LTR alignment of *A. lyrata* and *B. stricta HATE* homologs. Identified HSF recognition sites are orange (HRE H1) and grey (HRE H2) colored. Additional putative HSF recognition sites of HRE "X" in *B. stricta* are colored in green.

In summary, HREs are partially conserved in other *Brassicaceae* species. Here it was shown that *COPIA37* is present in other *Brassicaceae* species while the putative HRE A was almost conserved. However, additional putative binding sites also appeared in other species indicating non minor frequency alleles. Homologs of the *AlHATE* family TEs were identified in *B. stricta*, *B. rapa* and *E. salsugineum.* However, HREs were only present in *A. lyrata* and *B. stricta*. *HATE* TEs in both

species have a common putative HRE which were absent in some *B. stricta* and broken in two *A.lyrata HATE* TEs. However, both feature a putative HRE exclusively conserved within particular species, indicating an independent evolution of heat responsiveness. In contrast to this, *COPIA20* were exclusively identified within *A. lyrata* and *A. thaliana*, whereby HRE were only found in *AlCOPIA20* since *AtCOPIA20* were truncated and appeared by low copy number.

3.7 Validation of transcriptional activity upon heat induction of TE homologs in *Brassicaceae* that contain HRE

Transcriptional analysis was performed to test heat responsiveness of *COPIA20*, *COPIA37* and *HATE* containing or lacking putative HREs in different *Brassicaceae*. Plants were grown on 1/2 MS media and heat stressed for 6 hours at 37°C. Subsequently, heat induced transcriptional activity of candidate homologs TEs were tested by RT-qPCR relative to untreated plants.

Here, transcriptional activity was tested in *At- and AlCOPIA20* (Figure 3-23). This revealed no significant transcript accumulation in *A. thaliana*. In contrast, *AlCOPIA20* transcripts were significantly (t-test; α = 0.05) increased after 6 hours of stress, demonstrating putative HRE mediated transcriptional activity of *AlCOPIA20,* while transcriptional activity in *AtONSEN,* which does not feature HRE, remained silent.

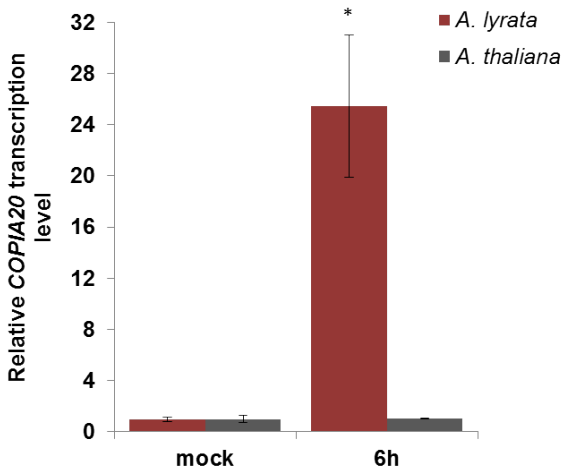

Figure 3-23: Relative transcript level of *COPIA20* relative to mock in *A. lyrata and A. thaliana* after 6 hours of heat treatment relative to mock. Expression was normalised to heat stress stable *GAPC-2* gene expression. Error bars indicate standard errors from two biological replicates. Statistical significance indicated by asterisks (t-test; α = 0.05).

Further, heat induced transcriptional activity was tested in *COPIA37* homologs in *A. lyrata, A. thaliana, C. rubella* and *E. salsugineum.* Significant (t-test; α = 0.05) transcriptional activity was observed in all tested *COPIA37* homologs after 6 hours of stress but *E. salsugineum* (Figure: 3-24). This indicated that only *COPIA37* TEs that contain the upstream inverted nGAAn-box motif (Figure: 3-21). As *EsCOPIA37* indicated being the only species that does not contain this motif, it can be assumed that this motif was mandatory for HSF recognition and binding to mediate transcriptional activity of *COPIA37*. Further, heat induced transcription was promoted in *AlCOPIA37*. As generic primers were used to determine transcriptional change, it can be assumed that heat response in *AlCOPIA37* is potentially promoted by the second identified HRE. However, this needs further investigations validating the HSF binding at *AlCOPIA37* HRE, which can be also mediated by alteration of epigenetic silencing.

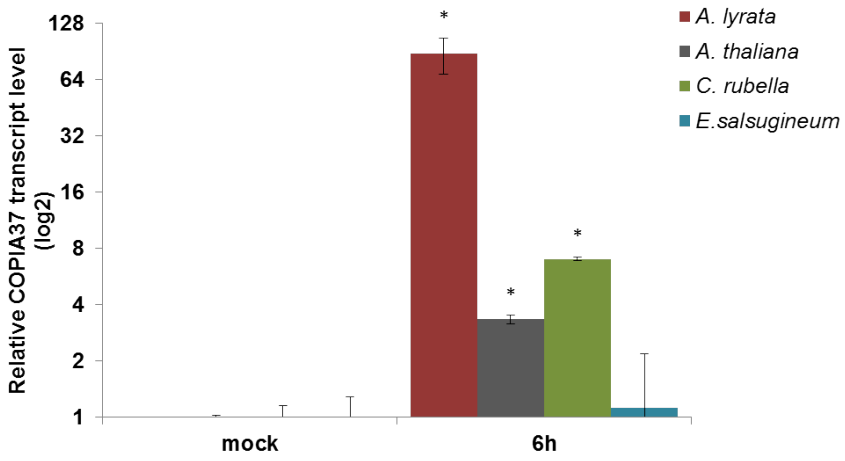

Figure 3-24: Relative transcript levels of *COPIA37* relative to mock in *A. lyrata, A. thaliana, C. rubella* and *E. salsugineum* after 6 and 12 hours of heat treatment. Normalised to *GAPC-2* gene expression. Error bars are standard errors from two biological replicates. Statistical significance indicated by asterisks (t-test; α = 0.05).

Heat induced transcriptional activity of *AlHATE* and *BsHATE*, indicated significant transcriptional increase (t-test; α = 0.05) of *HATE* transcripts within both species (Figure: 3-25). However, the transcript level of *BsHATE* was higher than that of *AlHATE.* As both feature one common HRE it can be assumed that maybe the second HRE exclusively identified in *BsHATE* promotes heat response. However, as delayed response to heat was shown also in *ONSEN* (3.2.1) and in whole transcriptome analysis, a broader heat tolerance of *A. lyrata* can be assumed.

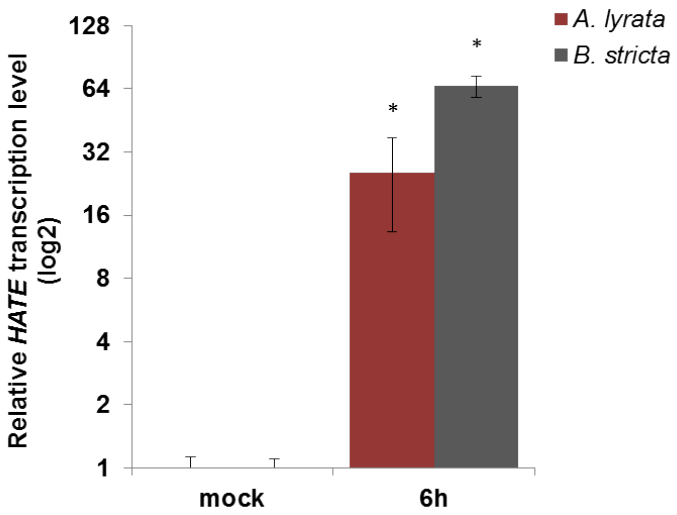

Figure 3-25: Relative transcript level of *HATE* relative to mock in *A. lyrata* and *B. stricta* after 6 and 12 hours of heat treatment. Normalised to *GAPC-2* gene expression. Error bars are standard errors from two biological replicates. Statistical significance indicated by asterisks (t-test; α = 0.05).

Summarized, *in vitro* analysis of heat induced transcriptional activity of TE families with novel identified HREs verified transcriptional response in all identified *COPIA37* and *HATE* homologs, except *EsCOPIA37* which is assumed by absence of a mandatory HSF-recognition site (nGAAn-box). This indicated conserved heat response of *COPIA37* and *HATE* in other *Brassicaceae*. In addition, *COPIA20* showed only heat induced activity in *A. lyrata*, the only species *COPIA20* was investigated, featuring HRE within its LTR. However, further analysis is needed in order to verify potential HSF binding at putative novel HREs.

4 Discussion

TEs were discovered by Barbara McClintock in 1940s (McClintock, 1950). Since then, they were considered mostly as “junk” DNA or genomic parasites (Ohno, 1972; Orgel et al., 1980). Many TEs reach high copy numbers in the host genome, owing to their amplification abilities by specific mechanisms. However, there is growing evidence that TEs play an important role in genome evolution, gene regulation and participate in plant developmental control (Tenaillon et al., 2010; Lisch, 2013a).

Transposon mobility represents a risk for genome stability (Miura et al., 2001; Mari-Ordonez et al., 2013). In order to prevent unwanted sequence changes, TEs are epigenetically suppressed by transcriptional gene silencing (TGS). However, recent studies suggested that TGS can be at least transiently attenuated under heat stress conditions, leading to transcriptional activity of specific TEs (Pecinka et al., 2010; Tittel-Elmer et al., 2010). Heat induced transcriptional activity of *COPIA78* family retrotransposons named *ONSEN* was shown in *A. thaliana* and even their transposition in mutants deficient in siRNA biogenesis (Pecinka et al., 2010; Ito et al., 2011). Recent publication postulated that *ONSEN* heat responsiveness is facilitated by the presence of HREs which reside in the *ONSEN* LTRs, and is mediated within the heat shock transcription factor-pathway (Cavrak et al., 2014). These findings demonstrated a potential mechanism for TEs to survive within the genome.

The aim of this thesis was to reconstruct the evolution of heat responsiveness within *Brassicaceae.* Therefore, whole transcriptome analysis was performed in the closely related species *A. thaliana* and *A. lyrata*, in order to identify common or novel heat responsive TEs. Further, heat responsive TEs were observed for common heat responsive motifs within their LTR region. Subsequently, homologs of heat responsive TEs were recovered in *Brassicaceae* species and observed for the presence of HRE.

4.1 Evolutionary conserved heat responsiveness among *COPIA* family TEs in *A. thaliana* and *A. lyrata*

Whole transcriptome analysis was performed in order to find common patterns and traits of TE heat responsiveness in the close related species *A. thaliana* and *A. lyrata*. The analysis revealed an overrepresentation of LTR-retrotransposons in response to heat relative to their overall genomic abundance in both species. This indicated promoted heat responsiveness within LTR-retrotransposons. This is not surprising, as stress responsive transcriptional activity was predominantly shown in LTR-retrotransposons (Casacuberta et al., 1997; Mhiri et al., 1997; Beguiristain et al., 2001; Ivashuta et al., 2002; Pecinka et al., 2010; Tittel-Elmer et al., 2010; Woodrow et al., 2010; Woodrow et al., 2011; Butelli et al., 2012; Grandbastien, 2015)

In plants, LTR-retrotransposons are divided into two superfamilies: *COPIA* and *Gypsy* (Wicker et al., 2007). Direct comparison of the fold-difference of TE abundance in response to heat relative to their genome wide spectrum revealed highest and almost equal abundance among *COPIA* elements in *A. thaliana* and *A. lyrata*. This finding suggested that stress-induced transcriptional activity is potentially widespread among LTR-retrotransposons, particularly within the *COPIA* superfamily. Similar findings were observed in adult Drosophila. Environmental stresses that induce heat shock response in adult Drosophila simultaneously activated *COPIA* LTR-retrotransposons (Strand et al., 1985). A recent publication indicated heat shock factor (HSF)-mediated expression of *COPIA* retrotransposon *ONSEN* in response to high-light stress as this share the same signaling pathway with heat stress (Matsunaga et al., 2015). Therefore it can be proposed that stress mediated expression is evolutionary conserved within *COPIA* LTR-retrotransposons. Although *COPIA*-elements were observed to mediate genic stress response in gene flanking regions (White et al., 1994; Kidwell et al., 2000; Ito et al., 2011), current knowledge does not support this as a purpose of TE transposition. The primary function of a TE is the survival within the genome (Doolittle et al., 1980; Hickey, 1982), which can be established by exploiting stress induced activity as a potential survival mechanism.

4.2 Conserved heat response of *ONSEN* in *A. thaliana* and *A. lyrata*

In order to test, whether the heat responsiveness could be evolutionary conserved, significantly upregulated *COPIA* TEs revealed by whole transcriptome sequencing in *A. thaliana* and *A. lyrata* were analyzed for common heat responsive TEs. If conservation was true, it would be expected that both species share a common set of heat responsive *COPIA* family TEs. Detailed analysis of heat responsive *COPIA* superfamily TEs in *A. thaliana* and *A. lyrata* showed an overrepresentation of *ONSEN* in both species. Although, heat induced increase of transcript level of *ONSEN in A. lyrata* was previously reported by RT-PCR (Ito et al., 2013), within this study particular heat responsive *ONSEN* elements could be identified and became available for further analysis. In total, 55 *ONSEN* elements were identified in *A. lyrata.* Thereof, 10 are full-length elements, while 45 are incomplete *ONSEN* elements including solo LTRs and truncated TEs. Ito et al. (2013) identified 17 *ONSEN* elements in *A. lyrata* whereby 12 were also identified within this study. However, identification of five *ONSEN* elements was not reproducible by Ito's study due to unpublished genomic location.

Reconstruction of *AlONSEN* consensus sequence out of all 10 full-length elements, indicated a slightly greater length compared to *AtONSEN* (5164 versus 4965 bp, respectively). The same holds true for the LTRs (440 bp versus 472 bp, respectively). Although the size of *ONSEN* in both species fits the average size of *COPIA* elements (Jurka, 2005), this size difference in *A. thaliana* correlates with the reported accumulation of small deletions that occur potentially during deletion-biased repair of DNA strand breaks (Hu et al., 2011).

Among all 55 *AlONSEN*, 19 showed transcriptional response to heat, including 9 previously identified *ONSENs* by Ito et al. (2013). With this, 34.5 % of *ONSEN* TEs showed transcriptional activity induced by heat. This is relatively the same amount as in *A. thaliana*, where there 33.3 % of *ONSEN* TEs are heat responsive. Besides a strong conservation of heat responsiveness by *ONSEN* TEs, this led to the assumption that heat induced activity of *ONSEN* TEs could be controlled by the demographic history. Similar findings were observed by *EVADÉ*/*AtCOPIA93* in *METHYLTRANSFERASE 1* (*met1)* epigenetic recombinant inbred lines (epiRILS). *EVADÉ* silencing occurred in generations with a copy number > 40 which coincided with LTR-derived 24-nt siRNA production (Mari-Ordonez et al., 2013). *ONSEN* elude TGS by HSF mediated transcriptional activity (Cavrak et al., 2014), recombination

potentially counteracts growing *ONSEN* copy number by truncation and elimination of *ONSEN* TEs. However, efficacy of selection against TEs is proposed as a function of location of a TE insertion (Tenaillon et al., 2010). TE insertions within exons could interrupt gene function, but were rarely investigated (Bartolomé et al., 2002; Rizzon et al., 2003; Wright et al., 2003; Tenaillon et al., 2010). In contrast to previous findings by semi-quantitative analysis (Ito et al., 2013), significantly transcriptionally upregulated *ONSEN* TEs were identified within this study. This not only validated conserved heat responsiveness among *ONSEN* TEs at relatively equal extent, it also revealed candidates to investigate for common traits that promote or mediate heat response.

4.3 Comparative analysis of *ONSEN* HRE in *A. lyrata* and *A. thaliana*

In order to investigate the evolution of heat responsiveness in *ONSEN,* elements were screened for a common trait. As in *A. thaliana*, *ONSEN* was reported to exploit heat stress for its activation by heat shock transcription factor (HSF) binding motifs within the LTRs (Cavrak et al., 2014). This heat responsive DNA element (HRE) is constituted by two palindromic repeats (nTTCnnGAAn or nGAAnnTTCn) and is recognized by a helix-turn-helix motif in the DNA binding domain (DBD) of a HSF (Littlefield et al., 1999; Cavrak, 2014). Thereby, at least three nGAAn-boxes are needed for high-affinity HSF recognition (Wu, 1995). However, *ONSEN* HREs consist of four nGAAn-boxes represented by two palindromic repeats, which constitute two possibilities for high affinity HSF-binding sites (A-and B-type motif). Within this study, such heat responsive DNA binding motifs (HRE) were identified in 27 out of the 55 *AlONSEN*, whereof 19 out of this 27 were identified in heat responsive *AlONSEN*. The remaining 8 *AlONSENs* were truncated TEs or solo LTRs*.* The identified HSF-binding sites of these truncated *AlONSEN* are basically remnant single palindromic repeats (nTTCnnGAAn), which cannot establish high-affinity HSF recognition (Wu, 1995; Sakurai et al., 2010).

The comparative analysis in this thesis could further show that among the 19 heat responsive *AlONSEN,* 8 feature the high affinity A- and B-type HSF binding sites. This includes six full length *AlONSEN* and two solo LTRs. In addition, a novel putative HSF binding motif was identified substituting the B-type, called C-type motif. The C-type motif evolved by a single base substitution $T_{145}C$ in the gap which links

the two palindromic repeats. Although this insertion interrupts the palindromic pattern, leaving a gap of 2 bp between the two nGAAn-boxes, HSF binding cannot be excluded as HSF can bind to a gap of at least 5 bp (Wu, 1995; Sakurai et al., 2010). Thus, it can be concluded that in this thesis a putative variant of the HSF-binding motif was identified. Besides this C-type motif, two *AlONSEN (AlONSEN8 & 13)* were missing the first palindromic repeat. This also means that an A-type HSF-binding site was undetectable. However, HSF-binding within these two *AlONSEN* can be maintained by an additional nTTCn motif 5 bp further downstream, that changes the perfect-type HRE to a gap-type HRE (Wu, 1995; Sakurai et al., 2010). Such variances were not observed exclusively in solo LTRs within this study. The C-type variance was identified in 2 full-length *AlONSEN*. In addition, one full-length *AlONSEN*, which lost the first palindromic repeat, indicated still possible heat inducibility and potential transposition activity of *AlONSEN* with HRE variances. Additionally, the transcriptional heat response of *AlONSEN* solo LTRs indicate that intra-strand homologous recombination, which leads to solo LTRs, is potentially not efficient because the remaining HRE still promote the accumulation of transcripts (Devos et al., 2002).

An LTR-based phylogenetic tree was reconstructed by using *ONSEN*-LTRs of *A. thaliana* and *A. lyrata* to investigate the evolution of HRE in both species. In addition, approximate ages of full-length *ONSEN* were estimated. The phylogenetic tree demonstrated two remarkable splits. *ONSEN*-LTRs of species diverged with a high bootstrap probability. This validated proposed species specific sequence variations in *ONSEN* (Ito et al., 2013). In addition, this confirmed that variations like in *AtONSEN4,* which led to a loss of the B-type HSF binding motif, and C-type in *AlONSEN* were established after species split as they do not occur in the other species.

The second split indicated a clear divergence of *AlONSEN* containing A-B-type HRE binding motifs and that of variant binding motifs. In addition, LTR age estimation revealed an overall younger TE age in *ONSEN* containing the C-type variance within its LTR than *ONSEN* containing A-and-B-type HSF-binding motifs. An exception are the full-length *AlONSEN11* and *15* whose HREs consist of A- and B-type HSF binding motifs and a low LTR divergence, indicating more recent transposition. These findings propose (a) that A-and B-type HSF binding motifs, which are almost conserved in *AtONSEN,* are potentially the ancestral HRE constitution and (b) that both A- and B- type as well as the variant HRE containing *ONSEN* have recently

transposed. In addition, these findings revealed a novel HSF binding site in *A. lyrata ONSEN* that mediates heat induced activity.

4.4 Evolution of HRE in *ONSEN*-LTRs

In order to reconstruct the evolution of HREs within *Brassicaceae,* the conservation of palindromic repeats was determined by phylogenetic shadowing. *ONSEN* TEs from different *Brassicaceae* species were identified by BLAST searches within Phytozome V.10 database. Thereby, a much lower number of *ONSEN* elements were identified in other *Brassicaceae*, while no *ONSEN* were detected in the *Capsella* genus.

Subsequently, reconstruction of LTRs and a search for HRE in identified *ONSEN* already indicated evolutionary conservation of one palindromic repeat. A third nGAAn-box for binding was observed in specific *ONSEN* in *A. alpina* and *B. rapa,* but indicated no positional conservation. However, HREs that constituted of two palindromic repeats, harbor four nGAAn HSF recognition sites, like in *A. thaliana* and *A. lyrata ONSEN* were identified exclusively in *B. antipoda*.

Here, the conservation of HREs were investigated, which should help to reconstruct the HRE evolution. Therefore, consensus sequences of LTRs containing HRE per species were via phylogenetic shadowing. This showed 100 % sequence conservation of the first palindromic repeat. However, the single palindromic repeat identified in *B. rapa* and *B. stricta* did not clearly align at the first or second palindromic repeat because of ambiguity within the AVID alignment. The single palindromic repeat was aligned to the position with fewer gaps. Strong conservation of the second palindromic repeat was only observed in species with two palindromic repeats (*B. antipoda* and *A. thaliana*). These data were subsequently used to propose a potential evolution of the perfect high-affinity type HRE as described (Wu, 1995; Sakurai et al., 2010).

The integration of the HRE conservation data into a diagramed phylogenetic tree made of established *Brassicaceae* phylogeny (Mandáková et al., 2010; Franzke et al., 2011; Heenan et al., 2012) proposes the presence of a single palindromic repeat in the ancestral *ONSEN*-LTR that underwent a local duplication leading to the high-affinity HRE within *Arabidopsis* and *Ballantinia* species. This hypothesis is strengthened by the number of identified *ONSEN* elements in the other species.

Here, fewer *ONSEN* were observed within species with a single palindromic repeat. This led to the assumption, that less complex HRE do not mediate heat responsive expression as efficient as the perfect high-affinity HRE (Wu, 1995; Sakurai et al., 2010; Enoki et al.). In addition, it can be suggested that TEs with perfect high-affinity HREs transpose more often or remain longer in the genome, indicated by an older LTR-Age in *A.lyrata ONSEN*-LTRs containing A-&B-type motif. With this it can be assumed that a survival mechanism may have been established in *ONSEN* by HSF DNA binding motifs. The proposed lower efficiency of a single palindromic repeat, although a putative third nGAAn-box was present more downstream, cannot counteract recombination, which leads to higher amount of solo LTRs, truncated elements and overall fewer amount of *ONSEN* TEs in those species. However, as high rate *ONSEN* transpositions were only observed in mutants deficient in siRNA biogenesis (Ito et al., 2011) and potentially in callus tissue of wild-type *A. thaliana (Matsunaga et al., 2012)*. It needs to be investigated when and to what extent *ONSEN* can amplify within the genome of wild-type plants, as recent amplifications were observed in *A. thaliana* and *A. lyrata ONSEN*.

4.5 Repeated evolution of heat responsiveness among *COPIA* TEs in *A. lyrata*

More heat responsive TEs were identified by whole transcriptome analysis. As previous investigations revealed a common trait, mediating heat response of *ONSEN* by HREs, LTRs were reconstructed and observed for the presence of putative HRE.

Here a novel common heat responsive family was identified with *COPIA37*. There were 4 *COPIA37* solo LTRs identified in *A. thaliana* (*AtCOPIA37)* and 5 full-length *COPIA37* in *A. lyrata* (*AlCOPIA37*). In addition, *AlCOPIA37* indicated almost unchanged transcriptional abundance after two days of recovery. Thus, making particular *AlONSEN37* a putative candidate for heat induced transposition. In addition, with *COPIA20* and *HATE* two novel heat responsive families were identified within *A. lyrata.* Here 9 *COPIA20* TEs and 6 *HATE* were identified. Subsequently, newly identified heat inducible *COPIA* family TEs were tested for putative HRE within their LTR, as this was identified as a common trait of heat responsiveness in *ONSEN*. There, different putative HRE were identified within the LTRs of each COPIA family. The HRE element found within *COPIA20*-LTR was constituted by an *ONSEN*-like palindromic repeat and a third nGAAn-box 3 bp upstream. These were

identified in six out of nine heat activated *AlCOPIA20,* but not in any of the residual 260 non-heat responsive *AlCOPIA20*.

Further, two putative HREs were observed in *COPIA37* and *HATE*. However, they completely differed from that in *ONSEN by* having at least 3 bp gaps between the nGAAn-boxes. In addition, the first putative *COPIA37* HRE was observed in two out of 57 *COPIA37* in *A. lyrata* (*AlCOPIA37*) and in four out of 32 *A. thaliana COPIA37* (*AtCOPIA37*), while the second motif was exclusively present in the *AlCOPIA37* with prolonged heat induced activity. Here, no non-heat responsive *AlCOPIA37* resided the HRE within their LTR. Surprisingly, in *A. thaliana* the HRE was only present in four non-heat responsive *AtCOPIA37.*

The two putative HRE in HATE were conserved in all six *A. lyrata HATE* TEs (*AlHATE*). As putative HRE led to the assumption of a simultaneous evolution of HRE within *COPIA* elements, such evolution was observed by a phylogenetic network of *A. lyrata* LTRs of heat active *COPIA20*, *37*, *HATE* and *ONSEN*. This revealed a repeated and independent evolution of heat responsiveness in *COPIA20*, *COPIA37*, *HATE* and *ONSEN*, indicated by an early divergence and large incompatible splits of *COPIA20*, *COPIA37* from *HATE* and *ONSEN*. However, although *ONSEN* and *HATE* are clearly distinct TE families, they are diverged by compatible split, which assumes a later divergence. Because of the fact, that both species differ in the composition of their HRE among other family specific sequence differences, it can be assumed that the split happened along the establishment of different HRE. The assumption that the heat responsiveness evolved repeatedly within particular species got underlined by previous findings of stress responsive TEs in other species e.g. cold inducible *Rider in Citrus sinensis* (Butelli et al., 2012) or wounding and freezing inducible *Tnt1 in Nicotiana tabacum* (Casacuberta et al., 1997; Mhiri et al., 1997; Beguiristain et al., 2001). Here, stress specific DNA binding motifs were identified within their 5´LTR, triggering the expression of these LTR-retrotransposons (Casacuberta et al., 2013). However, these elements were only identified within particular species and not repeatedly in any others.

This assumption was emphasized as novel heat responsive *COPIA* elements were observed in other species. Here, no *COPIA20* elements in any other genus than *Arabidopsis* were identified. However, no HRE were detected as well as any heat induced transcription in all 9 *AtCOPIA20*. This was verified by subsequent RT-qPCR

analysis on *A. thaliana* and *A. lyrata COPIA20* which showed no transcriptional increase after 6 hours of heat treatment in *AtCOPIA20.*

In contrast, *COPIA37* TEs were identified in *A. thaliana*, *B. rapa*, *C. rubella*, *C. grandiflora* and *E. salsugineum*. No putative HREs were identified in *B. rapa-COPIA37*. Two putative HREs were found in *A. lyrata COPIA37*, and only one was identified almost exclusively in solo LTRs of *COPIA37* in *A. thaliana*, *C. rubella*, *C. grandiflora* and *E. salsugineum*. However, one full-length *COPIA37* TE in *C. rubella* featured that single HRE. While putative HREs were identified in all *COPIA37 of C. rubella, C. grandiflora and E. salsugineum*, in *A. thaliana* putative HREs were identified in four out of 32 *AtCOPIA37* TEs. Here, heat induced transcription was also tested by RT-qPCR on stressed relative to unstressed plants. Heat induced expression of *COPIA37* was increased after 6 hours of heat in all species except *E. salsugineum*. *AlCOPIA37* transcription was somehow promoted compared to the other species, which could be caused by the second identified HRE in one *AlCOPIA37.* As generic primers were used, that are annealing to more than one TE copy, it was assumed, that the second HRE highly promotes heat response and in particular *AlCOPIA37* led to promoted and prolonged heat induced activity. This can be explained by the HRE constitution. As the first HRE is a step-type HRE, containing gaps of 2 bp between nGAAn boxes (nGAAn[2bp]nTTCn[2bp]nGAAn), the second HRE is a gap-type HRE, containing only one gap of up to 5 bp between two nGAAn-boxes (nTTCn[2bp]nGAAnnTTCn). Thereby, gap-type HRE mediate expression at high levels, while step-type HRE mediate expression on a basal or low level (Sakurai et al., 2010; Enoki et al., 2011). However, the correlation of the conserved putative HRE and the heat response of *COPIA37* remained unknown, as no *AtCOPIA37* element was up-regulated in whole transcriptome sequencing that contains a HRE, while those who reside a HRE within their LTR were not significantly upregulated in whole transcriptome sequencing. Therefore, it needs to be further tested if heat response in *COPIA37* is facilitated by alteration of epigenetic silencing, mediated by a HRE or their combination.

HATE elements were identified in *A. lyrata* (n = 6)*, B. stricta* (n = 14)*, B. rapa* (n = 2) *and E. salsugineum* (n = 6) (Appendix Table: A6). However, in *B. stricta HATE* were highly abundant in contrast to *A. lyrata*, as in total 14 *HATE* TEs were detected in *B. stricta*. Here, HREs were identified in 12 *BsHATE* and in all 6 *AlHATE* exclusively and not in *BrHATE* or *EsHATE.* In addition, *BrHATE* and *EsHATE* were almost all

truncated TEs and solo LTRs, except a single putative full-length *HATE* in *E. salsugineum*.

The first HRE (nTTCnnnnGAATTC) was almost conserved between *BsHATE* and *AlHATE,* while an additional HRE was observed in particular species. In addition to the conserved HRE H1, a putative gap-type HRE (nTTCnnGAAn[6bp]nGAAnnTTCn) upstream of the first one resides in *BsHATE* and downstream of the first HRE in *AlHATE*. This indicated repeated establishment of particular HRE within the same *COPIA* family in different species. This verified that particular motifs evolve repeatedly within *COPIA* species. Heat induced expression was tested by RT-qPCR and showed common transcriptional increase in *Al*- and *BsHATE*. However, *HATE* transcription was promoted in *B. stricta*.

In summary, identified novel heat responsive TEs revealed a repeated evolution of heat responsiveness within *COPIA* TEs. This was mainly indicated by novel identified TE family *HATE* and by *COPIA20*. While full-length *COPIA20* was exclusively identified in *A. lyrata* as a full elements residing HRE within their LTR and being heat activated, all *COPIA20* in *A. thaliana* were truncated, missing LTRs and thereby a HRE. In addition, no other observed *Brassicaceae* species contained a *COPIA20* copy. Here it was revealed that 6 out of 269 *COPIA20* copies contain a HRE, while 9 were heat activated after 6 hours of stress. It can be proposed, that *COPIA20* exploits heat stress by HSF mediated transcription.

Newly identified TE family *HATE* appears only in low abundance in *A. lyrata* and is highly heat responsive. In contrast, *HATE* was more abundant in *B. stricta*, while copies without HRE were observed in that species. HATE in both species established additional HREs which are exclusively within that species and do not appear in the *HATE* of the other species. As both variants showed heat responsiveness, an independent repeated evolution of HRE can be proposed for *HATE* in both species using the response to heat as a survival mechanism. However, further investigations are needed to verify the HSF mediated expression and if those elements really exploit heat stress to amplify and survive in their host genome.

4.6 Concluding remarks and outlook

Within this study, conserved heat responsiveness among *COPIA* LTR-retrotransposons was found. Heat-induced transcriptional activity within *COPIA* superfamily TEs was dominated by *ONSEN* due to the presence of HRE motifs and relatively high copy number. Conservation of that HRE was observed in *A. thaliana* and *A. lyrata*, while species specific mutations were established after species split. However, variants within the HSF binding motifs in some HREs were observed in *A. lyrata*, which do not seem to harm HSF-binding efficiency, based on proposed models (Wu, 1995; Sakurai et al., 2010; Enoki et al., 2011), and heat-induced expression. On the contrary, recent transpositions can be assumed in *A. lyrata ONSEN* with A- & B-type HREs as well as those containing C-type HRE variant. Further, *ONSEN* elements from other *Brassicaceae* were used to reconstruct evolution of HREs. This demonstrated that perfect high-affinity HRE, consisting of two palindromic repeats harboring four HSF binding motifs, evolved by a local duplication of a single palindromic repeat. This duplication event is dated before the split of the genera *Arabidopsis* and the *Ballantinia*. This also explained the low *ONSEN* copy numbers in the other investigated *Brassicaceae* species. Hereby, with the establishment of complex HREs, *ONSEN* established a survival strategy to survive within the host genome.

In addition, novel (*COPIA20*) and new common (*COPIA37* and *HATE*) heat responsive TEs within different families were identified using whole transcriptome sequencing. Thereby, a common trait that potentially promotes heat responsiveness of *COPIA* TEs was identified by HRE motifs. These motifs were conserved within other particular *Brassicaceae* species. However, further functional analysis is needed to validate, if heat responsiveness is mediated by a HSF that binds to the putative HRE within the LTRs. As no mutants in *A. lyrata* MN47 background are currently available, LTR-GUS-fusion constructs cloned into *A. thaliana* Col-0 wild-type and mutants deficient in HSFs could indicate HSF mediated transcriptional activity as previously in O*NSEN.* Furthermore, this thesis raised the question about the extent of HSF mediated expression. Recent transpositions can be assumed by low LTR divergence, however no transpositions were observed in previous studies in wild-type plants, despite accumulation of extrachromosomal DNA copies (Ito et al., 2011). A genetic screen using transposon display method on a population of post-stress progenies could identify potential new insertions.

In addition, the findings of this thesis raised the question if similar transcription factor binding motifs were established in other TEs triggering the expression under other biotic or abiotic stresses like cold, drought, wounding or plant microbe infections. Recent publication revealed high-light stress-mediated *ONSEN* transcription by *HSFA2* (Matsunaga et al., 2015). As such motifs were already observed in other species and TEs e. g. cold-inducible *MCIRE* in *Medicago trunculata (Ivashuta et al., 2002),* they can be used for interspecific analysis focusing on the regulation and evolution of such stress responsive TEs.

References

Abdelsamad, A., and Pecinka, A. (2014). Pollen-Specific Activation of Arabidopsis Retrogenes Is Associated with Global Transcriptional Reprogramming. The Plant Cell **26,** 3299-3313.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology **215,** 403-410.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology **11,** R106.

Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data.

Bartolomé, C., Maside, X., and Charlesworth, B. (2002). On the Abundance and Distribution of Transposable Elements in the Genome of Drosophila melanogaster. Molecular Biology and Evolution **19,** 926-937.

Beguiristain, T., Grandbastien, M.-A., Puigdomènech, P., and Casacuberta, J.M. (2001). Three Tnt1 Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco. Consequences for Retrotransposon Control and Evolution in Plants. Plant Physiology **127,** 212-221.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proceedings of the National Academy of Sciences **107,** 18724-18728.

Böhmdorfer, G., Hofacker, I.L., Garber, K., Jelenic, S., Nizhynska, V., Hirochika, H., Stadler, P.F., and Bachmair, A. (2005). Unorthodox mRNA start site to extend the highly structured leader of retrotransposon Tto1 mRNA increases transposition rate. RNA **11,** 1181-1191.

Bray, N., Dubchak, I., and Pachter, L. (2003). AVID: A Global Alignment Program. Genome Research **13,** 97-102.

Bryant, D., and Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Molecular Biology and Evolution **21,** 255-265.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., and Martin, C. (2012). Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. The Plant Cell **24,** 1242-1255.

Cappello, J., Handelsman, K., and Lodish, H. (1985). Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell **43,** 105-115.

Casacuberta, E., and González, J. (2013). The impact of transposable elements in environmental adaptation. Molecular Ecology **22,** 1503-1517.

Casacuberta, J., Vernhettes, S., Audeon, C., and Grandbastien, M.-A. (1997). Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. In Evolution and Impact of Transposable Elements, P. Capy, ed (Springer Netherlands), pp. 109-117.

Casacuberta, J.M., and Santiago, N. (2003). Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene **311,** 1-11.

Cavrak, V.V. (2014). *Characterization of ONSEN, a Copia-type retrotransposon in Arabidopsis thaliana* (University of Vienna).

Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M., and Mittelsten Scheid, O. (2014). How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. PLoS Genet **10,** e1004115.

Chandler, V.L., and Stam, M. (2004). Chromatin conversations: mechanisms and implications of paramutation. Nat Rev Genet **5,** 532-544.

Clarke, S.M., Cristescu, S.M., Miersch, O., Harren, F.J.M., Wasternack, C., and Mur, L.A.J. (2009). Jasmonates act with salicylic acid to confer basal thermotolerance in Arabidopsis thaliana. New Phytologist **182,** 175-187.

Cook, J., Martin, J., Lewin, A., Sinden, R., and Tristem, M. (2000). Systematic screening of Anopheles mosquito genomes yields evidence for a major clade of Pao-like retrotransposons. Insect Mol. Biol. **9,** 109-117.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research **16,** 10881-10890.

Devos, K.M., Brown, J.K.M., and Bennetzen, J.L. (2002). Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in Arabidopsis. Genome Research **12,** 1075-1079.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nature Genet. **35,** 41-48.

Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature **284,** 601-603.

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,** 584-598.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32,** 1792-1797.

Enoki, Y., and Sakurai, H. (2011). Diversity in DNA recognition by heat shock transcription factors (HSFs) from model organisms. FEBS Letters **585,** 1293-1298.

Evgen'ev, M. (1997). Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc. Natl Acad. Sci. USA **94,** 196-201.

Evgen'ev, M., and Arkhipova, I. (2005). Penelope-like elements [mdash] a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenet. Genome Res. **110,** 510-521.

Fedoroff, N.V. (2012). Transposable Elements, Epigenetics, and Genome Evolution. Science **338,** 758-767.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution **39,** 783-791.

Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. Nat Rev Genet **3,** 329-341.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends in Genetics **5,** 103-107.

Frankel, A.D., and Young, J.A. (1998). HIV-1: fifteen proteins and an RNA. Ann. Rev. Biochem. **67,** 1-25.

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in Plant Science **16,** 108-116.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Research **32,** W273-W279.

Fujimoto, R., Kinoshita, Y., Kawabe, A., Kinoshita, T., Takashima, K., Nordborg, M., Nasrallah, M.E., Shimizu, K.K., Kudoh, H., and Kakutani, T. (2008). Evolution and Control of Imprinted <italic>FWA</italic> Genes in the Genus <italic>Arabidopsis</italic>. PLoS Genet **4,** e1000048.

Gall, J.G. (1981). Chromosome structure and the C-value paradox. The Journal of Cell Biology **91,** 3s-14s.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics **28,** 2678-2679.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40,** D1178-D1186.

Goodwin, T., and Poulter, R. (2004a). A new group of tyrosine recombinase-encoding retrotransposons. Mol. Biol. Evol. **21,** 746-759.

Goodwin, T.J.D., and Poulter, R.T.M. (2004b). A New Group of Tyrosine Recombinase-Encoding Retrotransposons. Molecular Biology and Evolution **21,** 746-759.

Goodwin, T.J.D., Butler, M.I., and Poulter, R.T.M. (2003). Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. Microbiology **149,** 3099-3109.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Research **38,** W695-W699.

Grandbastien, M.-A. (2015). LTR retrotransposons, handy hitchhikers of plant regulation and stress response. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1849,** 403-416.

Grandbastien, M.A., Audeon, C., Bonnivard, E., Casacuberta, J.M., Chalhoub, B., Costa, A.P.P., Le, Q.H., Melayah, D., Petit, M., Poncet, C., Tam, S.M., van Sluys, M.A., and Mhiri, C. (2005). Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenetic and Genome Research **110,** 229-241.

Gregory, T.R. (2005). Synergy between sequence and size in Large-scale genomics. Nat Rev Genet **6,** 699-708.

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J., and Bennett, M.D. (2007). Eukaryotic genome size databases. Nucleic Acids Research **35,** D332-D338.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41,** 95-98.

Havecker, E., Gao, X., and Voytas, D. (2004). The diversity of LTR retrotransposons. Genome Biology **5,** 225.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., and Wendel, J.F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Research **16,** 1252-1261.

He, F., Zhang, X., Hu, J.-Y., Turck, F., Dong, X., Goebel, U., Borevitz, J.O., and de Meaux, J. (2012). Widespread Interspecific Divergence in Cis-Regulation of Transposable Elements in the Arabidopsis Genus. Molecular Biology and Evolution **29,** 1081-1091.

Heenan, P.B., Goeke, D.F., Houliston, G.J., and Lysak, M.A. (2012). Phylogenetic analyses of ITS and rbcL DNA sequences for sixteen genera of Australian and New Zealand Brassicaceae result in the expansion of the tribe Microlepidieae. Taxon **61,** 970-979.

Hickey, D.A. (1982). Selfish DNA: A Sexually-Transmitted Nuclear Parasite. Genetics **101,** 519-531.

Hirochika, H. (1993). Activation of tobacco retrotransposons during tissue culture. The EMBO Journal **12,** 2521-2528.

Hollister, J.D., Smith, L.M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B.S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proceedings of the National Academy of Sciences **108,** 2322-2327.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottilar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F.X., Van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., and Guo, Y.-L. (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet **43,** 476-481.

Hua-Van, A., Le Rouzic, A., Boutin, T.S., Filée, J., and Capy, P. (2011). The struggle for life of the genome's selfish architects. Biology Direct **6,** 19-19.

Huson, D.H., and Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution **23,** 254-267.

Ito, H., Yoshida, T., Tsukahara, S., and Kawabe, A. (2013). Evolution of the ONSEN retrotransposon family activated upon heat stress in Brassicaceae. Gene **518,** 256-261.

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,** 115-119.

Ivashuta, S., Naumkina, M., Gau, M., Uchiyama, K., Isobe, S., Mizukami, Y., and Shimamoto, Y. (2002). Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (Medicago sativa). The Plant Journal **31,** 615-627.

Jurka, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. **110,** 462-467.

Kaessmann, H., Vinckenbosch, N., and Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet **10,** 19-31.

Kajikawa, M., and Okada, N. (2002). LINEs mobilize SINEs in the eel through a shared 3[prime] sequence. Cell **111,** 433-444.

Kalendar, R., Flavell, A.J., Ellis, T.H.N., Sjakste, T., Moisy, C., and Schulman, A.H. (2011). Analysis of plant diversity with retrotransposon-based molecular markers. Heredity **106,** 520-530.

Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003). Arabidopsis MET1 Cytosine Methyltransferase Mutants. Genetics **163,** 1109-1122.

Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences **98,** 8714-8719.

Kapitonov, V.V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America **103,** 4540-4545.

Kawase, M., Fukunaga, K., and Kato, K. (2005). Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. Mol Genet Genomics **274,** 131-140.

Kidwell, M.G., and Lisch, D.R. (2000). Transposable elements and host genome evolution. Trends in Ecology & Evolution **15,** 95-99.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology **14,** R36.

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.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinformatics **23,** 1289-1291.

Kramerov, D., and Vassetzky, N. (2005). Short retroposons in eukaryotic genomes. Int. Rev. Cytol. **247,** 165-221.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., and Huala, E. (2011). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth **9,** 357-359.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics **23,** 2947-2948.

Larkindale, J., and Knight, M.R. (2002). Protection against Heat Stress-Induced Oxidative Damage in Arabidopsis Involves Calcium, Abscisic Acid, Ethylene, and Salicylic Acid. Plant Physiology **128,** 682-695.

Le, Q.H., Wright, S., Yu, Z., and Bureau, T. (2000). Transposon diversity in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **97,** 7376-7381.

Lisch, D. (2009). Epigenetic Regulation of Transposable Elements in Plants. Annual Review of Plant Biology **60,** 43-66.

Lisch, D. (2013a). How important are transposons for plant evolution? Nat Rev Genet **14,** 49-61.

Lisch, D.R. (2013b). Transposons in Plant Gene Regulation. In Plant Transposons and Genome Dynamics in Evolution (Wiley-Blackwell), pp. 93-116.

Littlefield, O., and Nelson, H.C.M. (1999). A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocrystal. Nat Struct Mol Biol **6,** 464-470.

Lockton, S., Ross-Ibarra, J., and Gaut, B.S. (2008). Demography and weak selection drive patterns of transposable element diversity in natural populations of Arabidopsis lyrata. Proceedings of the National Academy of Sciences of the United States of America **105,** 13965-13970.

Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I. (2006). Mechanisms of chromosome number reduction in Arabidopsis thaliana and related Brassicaceae species. Proceedings of the National Academy of Sciences of the United States of America **103,** 5224-5229.

Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K., and Lysak, M.A. (2010). Fast Diploidization in Close Mesopolyploid Relatives of Arabidopsis. The Plant Cell **22,** 2277-2290.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S.-s., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, R.A., and Wing, R.A. (2000). Rice Transposable Elements: A Survey of 73,000 Sequence-Tagged-Connectors. Genome Research **10,** 982-990.

Mari-Ordonez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet **45,** 1029-1039.

Matsunaga, W., Kobayashi, A., Kato, A., and Ito, H. (2012). The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of ONSEN, a copia-like retrotransposon in Arabidopsis thaliana. Plant and Cell Physiology **53,** 824-833.

Matsunaga, W., Ohama, N., Tanabe, N., Masuta, Y., Masuda, S., Mitani, N., Yamaguchi-Shinozaki, K., Ma, J.F., Kato, A., and Ito, H. (2015). A small RNA mediated regulation of a stress-activated retrotransposon and the tissue specific transposition during the reproductive period in Arabidopsis. Frontiers in Plant Science **6,** 48.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J.M. (2009). RNA-mediated chromatin-based silencing in plants. Current Opinion in Cell Biology **21,** 367-376.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet **15,** 394-408.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics **16,** 1046-1047.

McClintock, B. (1950). The Origin and Behavior of Mutable Loci in Maize. Proceedings of the National Academy of Sciences of the United States of America **36,** 344-355.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Research **41,** W597-W600.

Mhiri, C., Morel, J.-B., Vernhettes, S., Casacuberta, J., Lucas, H., and Grandbastien, M.-A. (1997). The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol **33,** 257-266.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Research **41,** e121.

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.

Morgan, M.T. (2001). Transposable element number in mixed mating populations. Genetics Research **77,** 261-275.

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okumoto, Y., Tanisaka, T., and Wessler, S.R. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature **461,** 1130-1134.

Ohno, S. (1972). So much "junk" DNA in our genome. Brookhaven symposia in biology **23,** 366-370.

Orgel, L.E., and Crick, F.H.C. (1980). Selfish DNA: the ultimate parasite. Nature **284,** 604-607.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M. (2010). The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana. Science **327,** 92-94.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., and Scheid, O.M. (2010). Epigenetic Regulation of Repetitive Elements Is Attenuated by Prolonged Heat Stress in Arabidopsis. The Plant Cell Online **22,** 3118-3129.

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

Peterson, D.G., Schulze, S.R., Sciara, E.B., Lee, S.A., Bowers, J.E., Nagel, A., Jiang, N., Tibbitts, D.C., Wessler, S.R., and Paterson, A.H. (2002). Integration of Cot Analysis, DNA Cloning, and High-Throughput Sequencing Facilitates Genome Characterization and Gene Discovery. Genome Research **12,** 795-807.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D.S., Jackson, S., Wing, R.A., and Panaud, O. (2006). Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Research **16,** 1262-1269.

Piégu, B., Bire, S., Arensburger, P., and Bigot, Y. (2015). A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. Molecular Phylogenetics and Evolution **86,** 90-109.

Pikaard, C.S., and Mittelsten Scheid, O. (2014). Epigenetic Regulation in Plants. Cold Spring Harbor Perspectives in Biology **6**.

Pritham, E.J., Putliwala, T., and Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene **390,** 3-17.

R Core Team. (2013). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rausch, J.W., and Le Grice, S.F.J. (2004). 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. The International Journal of Biochemistry & Cell Biology **36,** 1752-1766.

Rawat, V., Abdelsamad, A., Pietzenuk, B., Seymour, D., Koenig, D., Weigel, D., Pecinka, A., and Schneeberger, K. (2015). Improving the Annotation of *Arabidopsis lyrata* Using RNA-seq Data PLoS ONE.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics **16,** 276-277.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., and Mittler, R. (2004). When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress. Plant Physiology **134,** 1683-1696.

Rizzon, C., Martin, E., Marais, G., Duret, L., Ségalat, L., and Biémont, C. (2003). Patterns of selection against transposons inferred from the distribution of Tc1, Tc3 and Tc5 insertions in the mut-7 line of the nematode Caenorhabditis elegans. Genetics **165,** 1127-1135.

Sabot, F., Picault, N., El-Baidouri, M., Llauro, C., Chaparro, C., Piegu, B., Roulin, A., Guiderdoni, E., Delabastide, M., McCombie, R., and Panaud, O. (2011). Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. The Plant Journal **66,** 241-246.

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

Sakurai, H., and Enoki, Y. (2010). Novel aspects of heat shock factors: DNA recognition, chromatin modulation and gene expression. FEBS Journal **277,** 4140-4149.

Sambrook, J. (2001). Molecular cloning : a laboratory manual / Joseph Sambrook, David W. Russell. (Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory).

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

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.-T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., and Wilson, R.K. (2009). The B73

Maize Genome: Complexity, Diversity, and Dynamics. Science **326,** 1112-1115.

Schramm, F., Ganguli, A., Kiehlmann, E., Englich, G., Walch, D., and von Koskull-Döring, P. (2006). The Heat Stress Transcription Factor HsfA2 Serves as a Regulatory Amplifier of a Subset of Genes in the Heat Stress Response in Arabidopsis. Plant Mol Biol **60,** 759-772.

Seelamgari, A. (2004). Role of viral regulatory and accessory proteins in HIV-1 replication. Front. Biosci. **9,** 2388-2413.

Slotte, T., Hazzouri, K.M., Agren, J.A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A.E., Escobar, J.S., Newman, L.K., Wang, W., Mandakova, T., Vello, E., Smith, L.M., Henz, S.R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., Andolfatto, P., Hu, T.T., Blanchette, M., Clark, R.M., Quesneville, H., Nordborg, M., Gaut, B.S., Lysak, M.A., Jenkins, J., Grimwood, J., Chapman, J., Prochnik, S., Shu, S., Rokhsar, D., Schmutz, J., Weigel, D., and Wright, S.I. (2013). The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nat Genet **45,** 831-835.

Smit, A.F.A., Hubley, R., and Green, P. (2013-2015). RepeatMasker Open-4.0.

Soppe, W.J.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J.M. (2000). The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. Molecular Cell **6,** 791-802.

Strand, D.J., and McDonald, J.F. (1985). Copia is transcriptionally responsive to environmental stress. Nucleic Acids Research **13,** 4401-4410.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution.

Tenaillon, M.I., Hollister, J.D., and Gaut, B.S. (2010). A triptych of the evolution of plant transposable elements. Trends in Plant Science **15,** 471-478.

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

The International *Brachypodium* Initative. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature **463,** 763-768.

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant, I. (2010). Stress-Induced Activation of Heterochromatic Transcription. PLoS Genet **6,** e1001175.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. Nucleic Acids Research **40,** e115.

van der Knaap, E., Sanyal, A., Jackson, S.A., and Tanksley, S.D. (2004). High-Resolution Fine Mapping and Fluorescence in Situ Hybridization Analysis of sun, a Locus Controlling Tomato Fruit Shape, Reveals a Region of the Tomato Genome Prone to DNA Rearrangements. Genetics **168,** 2127-2140.

Vitte, C., and Bennetzen, J.L. (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proceedings of the National Academy of Sciences of the United States of America **103,** 17638-17643.

Wei, F., Zhang, J., Zhou, S., He, R., Schaeffer, M., Collura, K., Kudrna, D., Faga, B.P., Wissotski, M., Golser, W., Rock, S.M., Graves, T.A., Fulton, R.S., Coe, E., Schnable, P.S., Schwartz, D.C., Ware, D., Clifton, S.W., Wilson,

R.K., and Wing, R.A. (2009). The Physical and Genetic Framework of the Maize B73 Genome. PLoS Genet **5,** e1000715.

Weigel, D., and Glazebrook, J. (2002). Arabidopsis: A Laboratory Manual. Cold Spring Harbor Laboratory Press. 2002. 354 pages. ISBN 0 87969 573 0. Genetics Research **80,** 77-77.

Wheeler, T.J., and Eddy, S.R. (2013). nhmmer: DNA homology search with profile HMMs. Bioinformatics **29,** 2487-2489.

White, S.E., Habera, L.F., and Wessler, S.R. (1994). Retrotransposons in the flanking regions of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. Proceedings of the National Academy of Sciences **91,** 11792-11796.

Wicker, T., Zimmermann, W., Perovic, D., Paterson, A.H., Ganal, M., Graner, A., and Stein, N. (2005). A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-elF4E locus: Recombination, rearrangements and repeats. Plant Journal **41,** 184-194.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., and Schulman, A.H. (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet **8,** 973-982.

Willing, E.-M., Rawat, V., Mandáková, T., Maumus, F., James, G.V., Nordström, K.J.V., Becker, C., Warthmann, N., Chica, C., Szarzynska, B., Zytnicki, M., Albani, M.C., Kiefer, C., Bergonzi, S., Castaings, L., Mateos, J.L., Berns, M.C., Bujdoso, N., Piofczyk, T., de Lorenzo, L., Barrero-Sicilia, C., Mateos, I., Piednoël, M., Hagmann, J., Chen-Min-Tao, R., Iglesias-Fernández, R., Schuster, S.C., Alonso-Blanco, C., Roudier, F., Carbonero, P., Paz-Ares, J., Davis, S.J., Pecinka, A., Quesneville, H., Colot, V., Lysak, M.A., Weigel, D., Coupland, G., and Schneeberger, K. (2015). Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. Nature Plants **1**.

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

Woodrow, P., Pontecorvo, G., Fantaccione, S., Fuggi, A., Kafantaris, I., Parisi, D., and Carillo, P. (2010). Polymorphism of a new Ty1-copia retrotransposon in durum wheat under salt and light stresses. Theor Appl Genet **121,** 311-322.

Wright, S., and Schoen, D. (2000). Transposon dynamics and the breeding system. In Transposable Elements and Genome Evolution, J. McDonald, ed (Springer Netherlands), pp. 139-148.

Wright, S.I., Lauga, B., and Charlesworth, D. (2002). Rates and Patterns of Molecular Evolution in Inbred and Outbred Arabidopsis. Molecular Biology and Evolution **19,** 1407-1420.

Wright, S.I., Agrawal, N., and Bureau, T.E. (2003). Effects of Recombination Rate and Gene Density on Transposable Element Distributions in Arabidopsis thaliana. Genome Research **13,** 1897-1903.

Wu, C. (1995). Heat Shock Transcription Factors: Structure and Regulation. Annual Review of Cell and Developmental Biology **11,** 441-469.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and van der Knaap, E. (2008). A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit. Science **319,** 1527-1530.

Xiong, Y., Burke, W., and Eickbush, T. (1993). Pao, a highly divergent retrotransposable element from Bombyx mori containing long terminal repeats with tandem copies of the putative R region. Nucleic Acids Res. **21,** 2117-2123.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research **35,** W265-W268.

Yang, R., Jarvis, D.J., Chen, H., Beilstein, M., Grimwood, J., Jenkins, J., Shu, S., Prochnik, S., Xin, M., Ma, C., Schmutz, J., Wing, R.A., Mitchell-Olds, T., Schumaker, K., and Wang, X. (2013). The reference genome of the halophytic plant Eutrema salsugineum. Frontiers in Plant Science **4**.

Yoshida, T., Ohama, N., Nakajima, J., Kidokoro, S., Mizoi, J., Nakashima, K., Maruyama, K., Kim, J.-M., Seki, M., Todaka, D., Osakabe, Y., Sakuma, Y., Schöffl, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). Arabidopsis HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression. Mol Genet Genomics **286,** 321-332.

Yu, C., Zhang, J., and Peterson, T. (2011). Genome Rearrangements in Maize Induced by Alternative Transposition of Reversed Ac/Ds Termini. Genetics **188,** 59-67.

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLoS ONE **9,** e78644.

Appendix

Table A1: List of Oligonucleotides used in this study. Oligonucleotides marked as universal anneal to corresponding locus in all species used in this study.

Species	Gene / TE	Name	Sequence	Purpose
Universal	*GAPC-2*	GAPC-2_F	ATCGGTCGTTTGGTTGCTAGAGT	qPCR
	GAPC-2	GAPC-2_R	ACAAAGTCAGCTCCAGCCTCA	qPCR
	UBC28	UBC28qF	TCCAGAAGGATCCTCCAACTTCCTGCAGT	qPCR
	UBC28	UBC28qR	ATGGTTACGAGAAAGACACCGCCTGAATA	qPCR
	COPIA20	C20_qF2	TACATGAAGCCACCACCGGGT	qPCR
	COPIA20	C20_qR3	TCATCTCCGGGAATGACAAGGTA	qPCR
	COPIA37	C37_LTRR1	AACCGCTACGTCTCGGGG	qPCR
	HATE	HATE_u_R1	CACCTTGAGACATGCTCAAATA	qPCR
A. thaliana	*HSP101*	HSP101qF	TGAGCTAGCTGTGAATGCAGGACATGCTC	qPCR
	HSP101	HSP101qR	ATCACTCTTTCAGCAGATTGAGCTGCGTT	qPCR
	COPIA78	COPIA78qF2	CGGTGCTCACAAAGAGCAACTATG	qPCR
	COPIA78	COPIA78qR3	ATCCTTGATAGATTAGACAGAGAGCT	qPCR
	COPIA37	AtC37_F2	AGCTTAACTACAGAAGGGAAGGA	qPCR
	COPIA37	AtC37_R2	CTCTCCAATCTCTCATTTTCTCG	qPCR
A. lyrata	*COPIA78*	AlCOPIA78qF3	ACAATGCTCACAAAGAGCAACTATG	qPCR
	COPIA78	COPIA78qR3	ATCCTTGATAGATTAGACAGAGAGCT	qPCR
	COPIA37	AlyC37qF2	GACGCTCTGACAACCAACCT	qPCR
	COPIA37	AlyC37qR2	AACGCAGCCGAAGCTAATC	qPCR
	HATE	HATE_F2	GTATTACGGTCTTGGGCTAGTG	qPCR
	HATE	HATE_R2	ACCAAGTATGACTCCATACATGAC	qPCR
E. salsugineum	*COPIA37*	EsC37-u_F1	ACAGGTGGGYCTTTAATGGGC	qPCR
B. stricta	*HATE*	HATE_F1	TCATGTATGGAGTCATACTTGGT	qPCR
C. rubella	*COPIA37*	CrC37_F1	GGAGGTAGGTGAGACAAGACA	qPCR

Table A2: Results of the multiple pairwise alignments to identify closest *AtONSEN3* (AT5G13205) homolog in *A. lyrata*. Table represents TOP13 hits using BLAST multiple pairwise alignment tool.

Query ID	Subject ID	Identity %	Bit score	Query start	Query end	Subject start	Subject end
AT5G13205	AlyONSEN15	93.79	6100	1	4071	474	4526
AT5G13205	AlyONSEN1	93.93	4471	1111	4071	1634	4594
AT5G13205	AlyONSEN1	94.85	1814	1	1164	475	1635
AT5G13205	AlyONSEN4	93.89	6137	1	4071	481	4549
AT5G13205	AlyONSEN5	93.89	6137	1	4071	481	4549
AT5G13205	AlyONSEN6	94.04	6172	1	4071	479	4549
AT5G13205	AlyONSEN7	93.73	4257	1233	4071	1754	4593
AT5G13205	AlyONSEN7	94.21	1864	1	1227	473	1689
AT5G13205	AlyONSEN9	93.99	6157	1	4071	473	4539
AT5G13205	AlyONSEN11	93.91	6128	1	4071	474	4526
AT5G13205	AlyONSEN12	94.23	6213	1	4071	475	4542
AT5G13205	AlyONSEN13	94.50	6274	1	4071	463	4530
AT5G13205	AlyONSEN14	94.40	4715	1001	4069	405	3473

Table A3: List of significantly upregulated *A. lyrata* TEs in response to heat.

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
6	19710193	19710448	ATENSPM1A	DNA/En-Spm	11.33	39.91	3.52	1.82	3.00E-02
1	14922493	14923391	ATHATN10	DNA/hAT	0.00	6.38	Inf	Inf	1.43E-02
5	4327772	4328165	ATHATN10	DNA/hAT	0.00	44.37	Inf	Inf	8.04E-07
2	10326663	10326912	ATHATN8	DNA/hAT	0.00	58.48	Inf	Inf	2.18E-05
5	1793545	1793783	ATHATN9	DNA/hAT	0.00	8.50	Inf	Inf	1.51E-03
7	6685029	6685099	ATHAT1	DNA/hAT-Ac	0.00	12.92	Inf	Inf	1.65E-05
6	21179243	21181209	ATHAT9	DNA/hAT-Ac	0.00	5.59	Inf	Inf	2.34E-02
1	24053829	24053953	ATHAT9	DNA/hAT-Ac	2.28	27.47	12.04	3.59	8.54E-07
6	254801	255818	TAG2	DNA/hAT-Ac	14.05	38.51	2.74	1.45	1.63E-02
2	852477	853018	ATDNAI26T9	DNA/MuDR	0.00	5.63	Inf	Inf	1.25E-02
4	4993269	4994244	ATMU1	DNA/MuDR	28.30	103.83	3.67	1.88	7.31E-06
5	4328790	4329290	ATMU13	DNA/MuDR	0.00	15.43	Inf	Inf	1.71E-04
3	228801	229108	ATMU6N1	DNA/MuDR	54.71	120.86	2.21	1.14	5.59E-03
8	6164413	6164884	ATMU9	DNA/MuDR	0.61	64.34	105.06	6.72	2.94E-15
3	14608387	14608786	ATN9_1	DNA/MuDR	1.75	70.01	39.93	5.32	1.21E-16
8	6164984	6165207	BRODYAGA1A	DNA/MuDR	0.00	20.04	Inf	Inf	1.04E-06
6	12121298	12121563	BRODYAGA2	DNA/MuDR	2.28	17.83	7.81	2.97	1.75E-02
7	7998697	8006697	Vandal18	DNA/MuDR	8.43	40.44	4.80	2.26	2.43E-05
4	18879289	18881927	Vandal18	DNA/MuDR	3.34	16.00	4.79	2.26	1.10E-02
4	18882213	18890066	Vandal18	DNA/MuDR	12.91	54.48	4.22	2.08	2.18E-05
2	3948825	3951463	Vandal18	DNA/MuDR	3.51	14.13	4.03	2.01	1.80E-02
2	2369814	2372528	Vandal18	DNA/MuDR	5.71	22.39	3.92	1.97	3.99E-03
6	22008615	22011253	Vandal18	DNA/MuDR	10.80	41.06	3.80	1.93	3.52E-04
4	2431675	2434313	Vandal18	DNA/MuDR	7.46	27.93	3.74	1.90	2.54E-03

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
7	7995773	7998411	Vandal18	DNA/MuDR	6.85	25.46	3.72	1.90	5.10E-03
4	2434599	2442264	Vandal18	DNA/MuDR	21.35	77.12	3.61	1.85	2.56E-05
6	22000392	22008329	Vandal18	DNA/MuDR	23.88	69.62	2.92	1.54	5.46E-04
2	11662030	11662496	Vandal2N1	DNA/MuDR	0.00	155.02	Inf	Inf	6.87E-33
3	23409275	23409546	Vandal2N1	DNA/MuDR	0.00	30.77	Inf	Inf	1.50E-10
2	1584065	1584306	Vandal2N1	DNA/MuDR	0.53	28.20	53.35	5.74	1.12E-04
1	23589207	23590290	Vandal3	DNA/MuDR	5.62	26.88	4.78	2.26	9.57E-04
4	4689883	4692092	Vandal3	DNA/MuDR	70.58	160.93	2.28	1.19	1.87E-03
4	13918995	13919348	Vandal5A	DNA/MuDR	32.98	72.72	2.21	1.14	1.92E-02
1	12468349	12469273	ATIS112A	DNA/PIF-Harbinger	0.00	4.88	Inf	Inf	3.42E-02
9	141299	142610	ATIS112A	DNA/PIF-Harbinger	0.00	6.04	Inf	Inf	2.26E-02
5	6659525	6660433	ATIS112A	DNA/PIF-Harbinger	0.53	23.09	43.68	5.45	9.12E-08
5	14196002	14196162	ATIS112A	DNA/PIF-Harbinger	1.14	15.00	13.15	3.72	4.49E-04
6	2974340	2975240	ATIS112A	DNA/PIF-Harbinger	26.97	238.51	8.84	3.14	1.19E-16
8	3508480	3509252	ATIS112A	DNA/PIF-Harbinger	17.67	133.37	7.55	2.92	8.84E-12
5	7028443	7029166	ATIS112A	DNA/PIF-Harbinger	47.95	206.86	4.31	2.11	1.69E-08
1	27694660	27695442	ATIS112A	DNA/PIF-Harbinger	38.99	145.99	3.74	1.90	1.55E-06
2	9768764	9769584	ATIS112A	DNA/PIF-Harbinger	23.79	54.10	2.27	1.18	2.84E-02

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
3	22178568	22179583	ATIS112A	DNA/PIF-Harbinger	124.37	268.90	2.16	1.11	5.21E-03
2	10149768	10150183	ATIS112A	DNA/PIF-Harbinger	120.90	242.54	2.01	1.00	1.07E-02
6	10629112	10629566	Harbinger	DNA/PIF-Harbinger	1.67	13.25	7.93	2.99	3.70E-02
6	22889496	22889854	Harbinger	DNA/PIF-Harbinger	28.19	95.08	3.37	1.75	2.95E-05
2	964101	964546	Harbinger	DNA/PIF-Harbinger	63.04	135.47	2.15	1.10	4.92E-03
1	6930506	6930771	ATRAN	DNA?	0.00	17.26	Inf	Inf	6.93E-07
1	21653585	21653725	ATLINE1_1	LINE/L1	0.00	7.46	Inf	Inf	2.94E-02
6	10731056	10731824	ATLINE1_1	LINE/L1	0.00	13.18	Inf	Inf	1.22E-05
1	27784616	27785017	ATLINE1_10	LINE/L1	0.00	4.42	Inf	Inf	3.54E-02
1	27786821	27789184	ATLINE1_10	LINE/L1	0.00	33.88	Inf	Inf	1.11E-11
1	18138380	18139898	ATLINE1_10	LINE/L1	0.53	28.00	52.96	5.73	4.75E-06
7	6973477	6974462	ATLINE1_11	LINE/L1	0.00	12.29	Inf	Inf	1.15E-03
7	9956854	9957375	ATLINE1_11	LINE/L1	18.62	157.28	8.45	3.08	6.23E-14
7	11682398	11683682	ATLINE1_12	LINE/L1	0.00	5.13	Inf	Inf	2.42E-02
1	15195130	15195581	ATLINE1_2	LINE/L1	0.00	6.88	Inf	Inf	4.09E-02
983	2	3347	ATLINE1_2	LINE/L1	17.37	76.47	4.40	2.14	3.41E-04
4	22298701	22299269	ATLINE1_3A	LINE/L1	0.00	6.54	Inf	Inf	1.14E-02
7	14847939	14848273	ATLINE1_3A	LINE/L1	0.00	32.17	Inf	Inf	3.67E-11
9	1382285	1382778	ATLINE1_3A	LINE/L1	0.00	7.71	Inf	Inf	8.31E-03
4	7143786	7144111	ATLINE1_3A	LINE/L1	0.61	8.50	13.89	3.80	9.07E-03
2	17979649	17979887	ATLINE1_4	LINE/L1	0.00	6.79	Inf	Inf	8.31E-03

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
7	12785299	12787123	ATLINE1_5	LINE/L1	16.75	44.15	2.64	1.40	4.46E-03
4	257750	259993	ATLINE1_6	LINE/L1	10.19	63.26	6.21	2.63	1.05E-07
9	2019	2278	ATLINE1_6	LINE/L1	8.77	26.47	3.02	1.59	1.25E-02
2	18513178	18514248	ATLINE1_6	LINE/L1	36.60	75.70	2.07	1.05	2.30E-02
6	13546122	13547749	ATLINE1_7	LINE/L1	46.00	169.88	3.69	1.88	8.72E-07
3	17426625	17430741	ATLINE1A	LINE/L1	0.00	71.26	Inf	Inf	1.44E-20
5	4328167	4328654	ATLINE1A	LINE/L1	0.00	132.91	Inf	Inf	2.04E-30
5	11667087	11671200	ATLINE1A	LINE/L1	0.00	13.63	Inf	Inf	1.15E-05
6	19676090	19680206	ATLINE1A	LINE/L1	0.00	9.00	Inf	Inf	7.82E-04
9	874523	876855	ATLINE1A	LINE/L1	0.00	13.68	Inf	Inf	6.82E-06
1	14355461	14356082	ATLINE1A	LINE/L1	6.23	43.22	6.93	2.79	6.53E-07
3	23055184	23055603	ATLINEII	LINE/L1	0.00	9.51	Inf	Inf	4.03E-04
1	32616615	32620511	ATLINEII	LINE/L1	1.22	9.92	8.10	3.02	1.67E-02
159	13259	13869	ATLINEII	LINE/L1	34.59	210.40	6.08	2.60	1.06E-11
6	13648252	13649851	TA11	LINE/L1	2.98	35.68	11.98	3.58	3.97E-08
5	20238212	20238587	TA11	LINE/L1	3.34	36.18	10.83	3.44	1.03E-07
2	7763333	7763742	TA11	LINE/L1	1.06	9.25	8.75	3.13	1.42E-02
6	13647020	13647792	TA11	LINE/L1	2.28	11.42	5.00	2.32	2.70E-02
5	1469527	1470608	TSCL	LINE?	0.00	7.55	Inf	Inf	3.06E-03
6	6671103	6671299	AlCOPIA2	LTR/Copia	36.88	184.21	5.00	2.32	4.86E-10
10	3525	8063	AlCOPIA20	LTR/Copia	0.00	5.84	5.84	2.54	1.66E-02
10	141350	143663	AlCOPIA20	LTR/Copia	0.00	7.09	7.09	2.83	3.23E-03
10	149249	153800	AlCOPIA20	LTR/Copia	0.00	14.84	14.84	3.89	4.52E-06
10	159963	161320	AlCOPIA20	LTR/Copia	0.00	9.09	9.09	3.18	2.41E-02
10	294152	299187	AlCOPIA20	LTR/Copia	0.00	11.33	11.33	3.50	5.72E-04
10	455857	459788	AlCOPIA20	LTR/Copia	0.00	11.17	11.17	3.48	1.39E-04

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
11	214101	218623	AlCOPIA20	LTR/Copia	0.00	23.92	23.92	4.58	3.85E-08
28	61453	66005	AlCOPIA20	LTR/Copia	0.00	4.42	4.42	2.14	3.54E-02
3	24168379	24172931	AlCOPIA20	LTR/Copia	0.00	6.09	6.09	2.61	1.20E-02
7	1382758	1386993	AlCOPIA3	LTR/Copia	0.00	63.48	63.48	5.99	2.09E-06
6	2049297	2053262	AlCOPIA31	LTR/Copia	0.00	11.72	11.72	3.55	4.13E-05
5	15361073	15365054	AlCOPIA31	LTR/Copia	202.52	948.24	4.68	2.23	5.23E-10
8	6545676	6547471	AlCOPIA31A	LTR/Copia	1.84	10.76	5.86	2.55	1.16E-02
3	24128553	24130918	AlCOPIA37	LTR/Copia	0.00	10.01	10.01	3.32	2.20E-04
7	7640274	7644088	AlCOPIA37	LTR/Copia	0.00	85.80	85.80	6.42	5.06E-24
8	22068138	22070098	AlCOPIA37	LTR/Copia	4.40	50.18	11.41	3.51	1.82E-09
7	20963365	20968153	AlCOPIA37	LTR/Copia	2.98	18.12	6.09	2.61	2.24E-03
7	24639790	24641218	AlCOPIA37	LTR/Copia	54.41	129.92	2.39	1.26	2.31E-03
6	6671305	6671647	AlCOPIA4	LTR/Copia	64.43	304.41	4.72	2.24	1.70E-10
6	3424372	3424556	AlCOPIA42	LTR/Copia	0.53	15.16	28.68	4.84	1.08E-02
1	25447360	25449935	HATE1	LTR/Copia	0.00	298.64	298.64	8.22	1.48E-45
4	15135433	15137933	HATE2	LTR/Copia	0.00	77.32	77.32	6.27	3.73E-22
5	9945065	9947640	HATE3	LTR/Copia	0.00	259.05	259.05	8.02	4.76E-43
7	8285793	8288369	HATE4	LTR/Copia	0.00	230.57	230.57	7.85	7.22E-41
8	6161052	6163595	HATE5	LTR/Copia	0.00	31.51	31.51	4.98	3.40E-11
8	15681408	15682687	HATE6	LTR/Copia	0.00	152.87	152.87	7.26	1.99E-33
7	24603891	24604056	AlCOPIA49	LTR/Copia	1.14	14.96	13.11	3.71	1.19E-03
6	9776455	9776736	AlCOPIA6	LTR/Copia	0.00	4.88	Inf	Inf	3.42E-02
2	11103831	11107542	AlCOPIA62	LTR/Copia	45.73	112.79	2.47	1.30	1.64E-03
6	6675388	6675669	AlCOPIA67	LTR/Copia	0.00	14.97	Inf	Inf	3.61E-03
5	9947641	9948004	AlCOPIA7	LTR/Copia	0.00	10.42	Inf	Inf	3.61E-04
1	24919206	24919585	AlONSEN2	LTR/Copia	0.00	19.47	Inf	Inf	9.12E-08

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
1007	42	1687	AlONSEN19	LTR/Copia	0.00	95.36	Inf	Inf	5.61E-25
1007	2411	3819	AlONSEN16	LTR/Copia	0.00	134.90	Inf	Inf	2.64E-31
2	1120762	1121196	AlONSEN3	LTR/Copia	0.00	30.64	Inf	Inf	2.76E-11
2	4268383	4273330	AlONSEN4	LTR/Copia	0.00	96.70	Inf	Inf	2.17E-25
2	12788861	12789817	AlONSEN5	LTR/Copia	0.00	39.49	Inf	Inf	4.06E-05
5	4327023	4327410	AlONSEN10	LTR/Copia	0.00	15.05	Inf	Inf	5.58E-06
6	22653438	22658398	AlONSEN12	LTR/Copia	0.00	42.74	Inf	Inf	1.28E-11
638	3577	4818	AlONSEN17	LTR/Copia	0.00	463.04	Inf	Inf	8.55E-54
7	8283764	8284127	AlONSEN18	LTR/Copia	0.00	8.66	Inf	Inf	1.21E-02
7	23781774	23786710	AlONSEN13	LTR/Copia	0.00	143.66	Inf	Inf	2.15E-32
3	14350695	14355704	AlONSEN7	LTR/Copia	0.61	887.85	1449.86	10.50	1.85E-61
3	13033504	13038475	AlONSEN6	LTR/Copia	0.61	324.03	529.13	9.05	3.03E-45
4	16304811	16309766	AlONSEN9	LTR/Copia	0.61	221.89	362.35	8.50	1.40E-38
3	23055604	23059108	AlONSEN8	LTR/Copia	0.61	180.16	294.19	8.20	1.33E-34
247	3711	8654	AlONSEN15	LTR/Copia	1.22	217.81	177.84	7.47	5.97E-37
5	9883484	9888427	AlONSEN11	LTR/Copia	1.22	186.06	151.92	7.25	9.35E-34
1	11269188	11273881	AlONSEN1	LTR/Copia	1.14	150.68	132.05	7.04	4.03E-30
8	15966326	15970199	AlONSEN14	LTR/Copia	6.07	378.62	62.41	5.96	4.30E-39
6	1975273	1979605	AlCOPIA79	LTR/Copia	301.73	842.95	2.79	1.48	1.06E-04
5	1744668	1748999	AlCOPIA79	LTR/Copia	380.32	987.36	2.60	1.38	5.55E-04
4	10968279	10971132	AlCOPIA82	LTR/Copia	7.54	29.30	3.88	1.96	2.23E-03
9	1154457	1155437	AlCOPIA90	LTR/Copia	4.40	35.13	7.99	3.00	2.74E-06
667	349	3893	AlCOPIA93	LTR/Copia	187.72	526.83	2.81	1.49	4.82E-05
3	24131871	24133662	AlCOPIA95	LTR/Copia	9.21	185.14	20.10	4.33	1.10E-23
4	20662423	20666571	ATRE1	LTR/Copia	0.53	41.90	79.25	6.31	2.68E-13
1	9565607	9565865	Copia-2_AT	LTR/Copia	0.00	19.06	Inf	Inf	6.37E-08

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
1	29300265	29300780	ENDOVIR1	LTR/Copia	1.14	9.21	8.07	3.01	2.30E-02
7	21970377	21970546	ATGP10	LTR/Gypsy	0.53	95.16	180.00	7.49	8.93E-12
4	10383108	10383953	ATGP11	LTR/Gypsy	0.00	7.96	Inf	Inf	5.44E-03
2	8233905	8239705	ATGP2	LTR/Gypsy	9.49	39.20	4.13	2.05	4.24E-02
1	27785940	27786149	ATGP2N	LTR/Gypsy	0.00	207.36	Inf	Inf	9.42E-09
5	1686771	1690361	ATGP2N	LTR/Gypsy	1.06	18.01	17.03	4.09	2.09E-05
1	24049143	24052750	ATGP3	LTR/Gypsy	0.00	4.17	Inf	Inf	4.95E-02
8	17668999	17669316	ATGP3	LTR/Gypsy	13.16	46.44	3.53	1.82	3.42E-04
1	24047227	24047642	ATGP3A	LTR/Gypsy	0.00	20.38	Inf	Inf	7.97E-08
6	13613811	13617400	ATGP3A	LTR/Gypsy	21.18	89.79	4.24	2.08	1.12E-06
1	31509068	31512890	ATGP3B	LTR/Gypsy	22.13	101.99	4.61	2.20	1.49E-07
5	19207250	19207608	ATGP3B	LTR/Gypsy	24.77	61.39	2.48	1.31	9.70E-03
1	24052784	24053031	ATGP9B	LTR/Gypsy	1.59	66.45	41.90	5.39	1.31E-16
1	25978064	25978249	ATGP9B	LTR/Gypsy	1.75	9.46	5.40	2.43	4.74E-02
1	25978997	25979269	ATGP9B	LTR/Gypsy	5.37	19.72	3.67	1.88	1.38E-02
4	10842613	10843022	ATGP9B	LTR/Gypsy	31.09	106.13	3.41	1.77	3.74E-04
5	11380311	11380861	ATGP9B	LTR/Gypsy	16.59	52.24	3.15	1.66	3.57E-04
51	34086	35296	Athila3	LTR/Gypsy	1.06	8.75	8.28	3.05	2.41E-02
5	12487947	12489086	Athila4C	LTR/Gypsy	0.00	7.80	Inf	Inf	2.23E-03
6	13618669	13624022	Athila4C	LTR/Gypsy	34.43	123.28	3.58	1.84	6.04E-06
6	11404443	11405163	Athila4D	LTR/Gypsy	0.00	8.46	Inf	Inf	2.75E-03
2	13152619	13153718	Athila4D	LTR/Gypsy	1.67	25.99	15.57	3.96	9.77E-03
5	6824106	6825724	Athila6B	LTR/Gypsy	0.61	15.75	25.72	4.69	3.75E-05
2	8240455	8242079	Atlantys1	LTR/Gypsy	2.20	18.46	8.40	3.07	3.01E-04
2	8243364	8244028	TAT1_ATH	LTR/Gypsy	0.00	7.00	Inf	Inf	1.10E-02
5	9950151	9951074	ATREP11A	RC/Helitron	0.00	186.65	Inf	Inf	5.58E-03

Scaffold	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
6	10891965	10894112	ATREP15	RC/Helitron	0.00	7.25	Inf	Inf	8.01E-03
1	27005598	27005900	ATREP4	RC/Helitron	0.00	11.38	Inf	Inf	1.82E-04
4	4544550	4545034	ATREP4	RC/Helitron	0.00	17.12	Inf	Inf	7.16E-05
7	21967975	21968105	ATREP5	RC/Helitron	0.00	26.58	Inf	Inf	2.00E-09
8	12849625	12850101	Helitron1	RC/Helitron	13.61	37.82	2.78	1.47	5.21E-03
6	20538407	20538883	Helitron1	RC/Helitron	18.87	50.28	2.66	1.41	2.71E-03
6	20539142	20540480	Helitron2	RC/Helitron	12.11	32.62	2.69	1.43	3.14E-02
8	12845477	12849366	Helitron2	RC/Helitron	16.06	36.47	2.27	1.18	4.22E-02
8	14373845	14374134	Helitron4	RC/Helitron	13.69	50.43	3.68	1.88	2.54E-04
6	3791332	3791481	Helitron5	RC/Helitron	0.00	5.33	Inf	Inf	3.32E-02
3	2021174	2021483	Helitron5	RC/Helitron	2.11	13.18	6.23	2.64	5.86E-03
1	14962699	14963246	HelitronY1B	RC/Helitron	0.00	8.96	Inf	Inf	1.43E-03
1	23719425	23719728	AtSB3	SINE	134.87	367.02	2.72	1.44	6.21E-05
3	19537075	19537364	Sadhu2-1	SINE	1.06	15.71	14.86	3.89	3.41E-04
4	1991313	1992053	Sadhu4-2	SINE	110.82	224.09	2.02	1.02	8.71E-03
2	2825662	2826492	Sadhu6-1	SINE	0.00	10.51	Inf	Inf	1.14E-04
5	13795510	13796337	Sadhu6-1	SINE	0.00	28.52	Inf	Inf	6.66E-11
4	5805364	5806193	Sadhu6-1	SINE	4.40	378.59	86.10	6.43	4.32E-21
5	13548481	13549298	Sadhu6-1	SINE	5.18	379.56	73.32	6.20	1.06E-40
5	4327520	4327648	Sadhu8-1	SINE	0.00	42.60	Inf	Inf	9.58E-15
8	3932951	3933117	AtSB2	SINE/tRNA	5.09	75.46	14.82	3.89	2.05E-11
7	7606754	7607100	AtSB6	SINE/tRNA	0.00	19.22	Inf	Inf	1.17E-07
2	12122982	12123724	ATSINE2A	SINE?	6.76	103.16	15.25	3.93	2.44E-08

Table A4: Significantly upregulated *A. thaliana* TEs in response to heat.

Chr.	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
4	3501968	3502431	ATDNA12T3A	DNA	0.00	23.47	Inf	Inf	2.00E-09
3	14240635	14241208	ATENSPM10	DNA/En-Spm	21.35	56.01	2.62	1.39	2.82E-02
1	13839342	13840830	ATENSPM12	DNA/En-Spm	0.32	637.65	1977.18	10.95	5.70E-60
4	2010968	2013986	ATENSPM2	DNA/En-Spm	0.00	54.47	Inf	Inf	2.07E-16
1	23742384	23742483	ATHAT1	DNA/hAT-Ac	5.55	24.03	4.33	2.11	1.87E-03
3	8978243	8978405	ATHATN1	DNA/hAT-Ac	0.00	5.87	Inf	Inf	2.85E-03
1	21746796	21747327	ATHATN3	DNA/hAT-Ac	29.90	147.87	4.95	2.31	4.52E-07
1	21524157	21524459	ATHATN5	DNA/hAT-Ac	0.00	52.31	Inf	Inf	8.04E-13
2	11081250	11081712	TAG2	DNA/hAT-Ac	0.00	10.87	Inf	Inf	5.93E-05
1	11315966	11316663	TAG2	DNA/hAT-Ac	0.32	13.95	43.26	5.43	2.89E-05
5	2870854	2871214	TAG2	DNA/hAT-Ac	1.51	13.36	8.84	3.14	2.03E-03
4	7426340	7426940	ArnoldY2	DNA/MuDR	0.00	41.10	Inf	Inf	2.01E-04
5	18405953	18406403	ATDNA1T9A	DNA/MuDR	150.55	430.97	2.86	1.52	9.60E-05
2	6881261	6884598	ATMU1	DNA/MuDR	37.22	179.77	4.83	2.27	9.68E-07
4	1872431	1872632	ATMU2	DNA/MuDR	0.32	5.31	16.46	4.04	3.32E-02
5	12966786	12968156	Vandal20	DNA/MuDR	38.23	197.87	5.18	2.37	1.17E-05
2	4729652	4731094	Vandal21	DNA/MuDR	0.32	216.28	670.62	9.39	4.48E-20
4	1612599	1613940	Vandal5	DNA/MuDR	16.12	48.01	2.98	1.57	7.04E-03
1	5030107	5030262	ATIS112A	DNA/PIF-Harbinger	0.00	10.00	Inf	Inf	3.75E-03
5	13004467	13009566	ATIS112A	DNA/PIF-Harbinger	0.00	5.87	Inf	Inf	2.85E-03
3	17577495	17578250	ATIS112A	DNA/PIF-Harbinger	185.10	523.10	2.83	1.50	2.22E-05
2	14143132	14143432	ATLINE1_11	LINE/L1	1.56	88.58	56.71	5.83	3.55E-13
1	23974197	23974493	ATLINE1_11	LINE/L1	5.88	18.70	3.18	1.67	4.06E-02

Chr.	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
5	22628447	22631434	ATLINE1_11	LINE/L1	458.56	1386.60	3.02	1.60	2.38E-06
5	17236714	17237162	ATLINE1_12	LINE/L1	0.97	35.39	36.58	5.19	8.08E-09
5	15383223	15383913	ATLINE1_3A	LINE/L1	11.16	35.93	3.22	1.69	2.11E-02
2	3355577	3356473	ATLINE1_6	LINE/L1	0.00	8.13	Inf	Inf	3.11E-02
5	9870902	9871248	ATLINE1_9	LINE/L1	0.00	8.08	Inf	Inf	1.04E-03
4	7061025	7062440	ATLINEII	LINE/L1	0.00	5.31	Inf	Inf	6.98E-03
1	10723463	10725219	ATLINEII	LINE/L1	6.05	40.27	6.66	2.74	2.19E-04
2	1922585	1925313	TA11	LINE/L1	0.59	8.92	15.01	3.91	1.91E-03
5	1165575	1167126	TA11	LINE/L1	38.06	117.15	3.08	1.62	5.75E-04
3	19085515	19085779	TSCL	LINE?	107.12	292.19	2.73	1.45	6.48E-03
5	20048358	20049583	TSCL	LINE?	1144.05	2105.08	1.84	0.88	6.85E-03
3	23100532	23104546	ATCopia23	LTR/Copia	26.63	70.81	2.66	1.41	1.38E-02
3	12607338	12608254	ATCopia28	LTR/Copia	0.32	37.16	115.23	6.85	1.04E-11
2	1388832	1393901	ATCopia28	LTR/Copia	12.43	301.94	24.30	4.60	2.45E-17
5	26415607	26415807	ATCopia32B	LTR/Copia	0.32	56.60	175.49	7.46	9.08E-15
4	1875175	1875724	ATCopia37	LTR/Copia	16.83	126.10	7.49	2.91	1.88E-10
3	10396358	10396888	ATCopia37	LTR/Copia	4.31	28.80	6.68	2.74	1.86E-04
1	3895125	3895672	ATCopia37	LTR/Copia	9.70	45.21	4.66	2.22	3.40E-04
3	10400031	10400574	ATCopia37	LTR/Copia	10.56	48.01	4.55	2.18	2.03E-04
2	3379194	3379771	ATCopia38A	LTR/Copia	0.00	3.92	Inf	Inf	3.48E-02
3	7368941	7369165	ATCopia39	LTR/Copia	0.32	5.87	18.21	4.19	1.53E-02
2	7598600	7602611	ATCopia39	LTR/Copia	1.19	15.00	12.61	3.66	1.45E-02
4	9498898	9499314	ATCopia49	LTR/Copia	0.00	32.21	Inf	Inf	5.60E-04
1	11178296	11178884	ATCopia50	LTR/Copia	0.00	12.31	Inf	Inf	4.12E-05
2	3301840	3303207	ATCopia51	LTR/Copia	2.16	79.23	36.73	5.20	2.57E-04
3	15768787	15773747	ATCopia62	LTR/Copia	366.87	682.52	1.86	0.90	3.28E-02
1	15609275	15609514	ATCopia66	LTR/Copia	0.00	7.80	Inf	Inf	1.63E-03

Chr.	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
4	2989895	2990052	ATCopia69	LTR/Copia	0.00	7.80	Inf	Inf	1.63E-03
1	7714707	7715145	COPIA78	LTR/Copia	0.00	147.18	Inf	Inf	9.82E-30
3	13369174	13374107	ATONSEN8	LTR/Copia	1.19	1572.24	1322.09	10.37	3.81E-84
1	3360540	3361316	COPIA78	LTR/Copia	1.19	653.68	549.68	9.10	3.96E-59
1	7717356	7722547	ATONSEN7	LTR/Copia	8.20	4291.20	523.09	9.03	8.61E-100
3	22059535	22064329	ATONSEN6	LTR/Copia	11.82	6141.95	519.55	9.02	7.29E-104
1	3780765	3785720	ATONSEN1	LTR/Copia	2.48	1270.94	512.63	9.00	1.09E-73
1	18013162	18018751	ATONSEN5	LTR/Copia	22.35	11241.88	502.92	8.97	6.40E-100
3	22695566	22700521	ATONSEN2	LTR/Copia	4.26	1876.45	440.17	8.78	1.21E-71
1	21524995	21529850	ATONSEN4	LTR/Copia	15.49	6619.06	427.31	8.74	3.84E-102
5	4208083	4213084	ATONSEN3	LTR/Copia	9.10	2358.18	259.12	8.02	1.14E-45
5	10018818	10020231	ATCopia95	LTR/Copia	1.88	16.69	8.86	3.15	2.45E-03
1	21796938	21801751	ATRE1	LTR/Copia	27.87	655.35	23.52	4.56	3.68E-32
1	21833196	21838009	ATRE1	LTR/Copia	5.98	74.78	12.51	3.65	5.04E-10
1	12813938	12816114	ROMANIAT5	LTR/Copia	0.00	6.72	Inf	Inf	1.91E-03
1	15610627	15615403	ROMANIAT5	LTR/Copia	0.00	31.34	Inf	Inf	8.25E-06
1	13230993	13235689	ROMANIAT5	LTR/Copia	0.59	103.79	174.55	7.45	1.77E-22
3	12603266	12606828	ROMANIAT5	LTR/Copia	0.32	51.49	159.66	7.32	3.51E-13
1	14481785	14481891	ATGP10	LTR/Gypsy	0.00	7.26	Inf	Inf	1.12E-03
3	13919450	13924434	ATGP10	LTR/Gypsy	0.00	5.59	Inf	Inf	4.49E-03
4	3859462	3860582	ATGP10	LTR/Gypsy	0.00	17.85	Inf	Inf	2.36E-07
5	11194184	11194399	ATGP10	LTR/Gypsy	0.00	12.26	Inf	Inf	2.65E-05
2	6884599	6885132	ATGP10	LTR/Gypsy	4.96	891.23	179.74	7.49	3.73E-53
2	2562810	2570724	ATGP2	LTR/Gypsy	0.00	3.92	Inf	Inf	3.48E-02
2	3294341	3295667	ATGP4	LTR/Gypsy	0.00	9.46	Inf	Inf	4.22E-04
4	15559246	15562581	ATGP9B	LTR/Gypsy	2.48	170.24	68.66	6.10	7.97E-26
5	4788501	4791446	ATGP9B	LTR/Gypsy	2.16	146.54	67.95	6.09	3.28E-24

Chr.	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
5	4799158	4799906	ATGP9B	LTR/Gypsy	54.14	524.66	9.69	3.28	1.64E-19
1	14248937	14255836	Athila2	LTR/Gypsy	0.00	9.46	Inf	Inf	4.22E-04
1	14334331	14341289	Athila2	LTR/Gypsy	0.00	5.56	Inf	Inf	1.66E-02
1	14425176	14426724	Athila2	LTR/Gypsy	0.00	15.95	Inf	Inf	1.17E-04
3	13405033	13412098	Athila2	LTR/Gypsy	0.00	14.52	Inf	Inf	1.66E-06
4	3087019	3099354	Athila2	LTR/Gypsy	0.00	20.11	Inf	Inf	2.31E-08
4	3537215	3547183	Athila2	LTR/Gypsy	0.00	20.85	Inf	Inf	3.24E-05
4	4768682	4777209	Athila2	LTR/Gypsy	0.00	24.11	Inf	Inf	6.34E-03
5	11796906	11807884	Athila2	LTR/Gypsy	0.00	75.66	Inf	Inf	6.51E-20
5	12128971	12138055	Athila2	LTR/Gypsy	0.00	14.80	Inf	Inf	1.21E-06
4	3173302	3184236	Athila2	LTR/Gypsy	1.19	82.01	68.97	6.11	2.99E-18
5	12649091	12650823	Athila2	LTR/Gypsy	0.32	6.98	21.63	4.43	9.44E-03
4	3344833	3354744	Athila2	LTR/Gypsy	0.32	5.31	16.46	4.04	3.32E-02
3	14788105	14789521	Athila3	LTR/Gypsy	1.56	51.32	32.85	5.04	1.64E-11
3	14476975	14477416	Athila4A	LTR/Gypsy	0.00	8.64	Inf	Inf	4.46E-04
2	4731304	4732460	Athila4C	LTR/Gypsy	0.00	76.11	Inf	Inf	1.96E-19
3	14477427	14478348	Athila4C	LTR/Gypsy	0.00	18.44	Inf	Inf	5.79E-08
4	4777208	4778759	Athila5	LTR/Gypsy	0.00	6.95	Inf	Inf	6.14E-03
1	15412724	15418821	Athila6A	LTR/Gypsy	0.00	3.64	Inf	Inf	3.23E-02
1	15705623	15715037	Athila6A	LTR/Gypsy	0.00	41.37	Inf	Inf	2.36E-11
2	3865229	3875676	Athila6A	LTR/Gypsy	0.00	9.23	Inf	Inf	6.59E-05
3	12245604	12255282	Athila6A	LTR/Gypsy	0.00	3.64	Inf	Inf	3.23E-02
3	13643236	13654839	Athila6A	LTR/Gypsy	0.00	5.87	Inf	Inf	2.85E-03
4	1804780	1811315	Athila6A	LTR/Gypsy	0.00	7.51	Inf	Inf	2.50E-03
4	4959998	4967794	Athila6A	LTR/Gypsy	0.00	3.64	Inf	Inf	3.23E-02
5	11320516	11331861	Athila6A	LTR/Gypsy	0.00	16.77	Inf	Inf	1.45E-07
5	11686010	11697351	Athila6A	LTR/Gypsy	0.00	13.41	Inf	Inf	2.38E-06

Chr.	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
5	11808470	11821681	Athila6A	LTR/Gypsy	0.00	17.85	Inf	Inf	2.36E-07
5	12139290	12150833	Athila6A	LTR/Gypsy	0.00	20.39	Inf	Inf	1.78E-08
5	12641362	12649090	Athila6A	LTR/Gypsy	0.00	21.41	Inf	Inf	1.74E-06
5	12996438	13004466	Athila6A	LTR/Gypsy	0.00	6.44	Inf	Inf	1.22E-03
5	13318996	13328540	Athila6A	LTR/Gypsy	0.00	20.11	Inf	Inf	2.31E-08
2	3636370	3645326	Athila6A	LTR/Gypsy	0.32	98.51	305.43	8.25	7.81E-22
4	3274435	3285235	Athila6A	LTR/Gypsy	0.92	48.44	52.82	5.72	8.58E-12
3	15713230	15721186	Athila6A	LTR/Gypsy	1.29	62.19	48.21	5.59	2.70E-13
4	4016073	4025224	Athila6A	LTR/Gypsy	0.32	10.36	32.13	5.01	8.15E-03
5	11830762	11851815	Athila6A	LTR/Gypsy	1.19	29.85	25.10	4.65	1.04E-08
3	14793856	14803440	Athila6A	LTR/Gypsy	2.11	27.06	12.85	3.68	1.02E-06
3	13630247	13637919	Athila6B	LTR/Gypsy	0.00	26.29	Inf	Inf	1.89E-07
1	15823077	15847509	ATHILA6C	LTR/Gypsy	0.00	11.67	Inf	Inf	2.66E-03
3	13902135	13913817	ATHILA6C	LTR/Gypsy	0.00	19.80	Inf	Inf	7.03E-08
4	3736029	3750029	ATHILA6C	LTR/Gypsy	0.00	4.46	Inf	Inf	2.90E-02
1	15178942	15191268	ATHILA6C	LTR/Gypsy	0.32	10.05	31.17	4.96	4.49E-04
2	5565096	5578989	ATHILA6C	LTR/Gypsy	0.32	5.59	17.34	4.12	2.28E-02
2	3308220	3310656	Athila7	LTR/Gypsy	0.59	17.90	30.11	4.91	1.64E-05
4	3503680	3505892	Atlantys2	LTR/Gypsy	0.92	134.97	147.17	7.20	4.86E-25
3	22513496	22521216	Atlantys2	LTR/Gypsy	37.04	230.92	6.23	2.64	3.49E-11
5	9906127	9907119	Atlantys2	LTR/Gypsy	3.35	19.77	5.91	2.56	1.63E-03
5	9907459	9908639	Atlantys2	LTR/Gypsy	3.57	13.64	3.82	1.94	2.11E-02
5	4794415	4796567	Atlantys3	LTR/Gypsy	0.65	79.32	122.98	6.94	2.40E-18
3	3111278	3115854	Atlantys3	LTR/Gypsy	1.88	13.13	6.97	2.80	3.45E-03
5	26415173	26415606	TAT1_ATH	LTR/Gypsy	0.92	170.62	186.05	7.54	1.02E-29
5	21556084	21556320	Helitron1	RC/Helitron	22.43	64.62	2.88	1.53	6.98E-03
3	12361779	12362224	Helitron3	RC/Helitron	3.35	13.92	4.16	2.06	2.72E-02

Chr.	Start	End	Family	Superfamily	BaseMean mock	BaseMean heat	Fold Change	Log2-FoldChange	Padj
5	11441830	11444761	Helitron5	RC/Helitron	0.00	13.92	Inf	Inf	8.44E-06

Table A5: List of identified *ONSEN* in *A. lyrata.* Heat responsive *AlONSEN* are indicated by numeration. *AlONSEN* previously identified by Ito et al. (2013) are enlisted with corresponding Locus-ID.

Name/Family	Scaff.	Start	End	HREs	Locus ID Ito et al. (2013)	Notes
AlONSEN1	1	11269188	11273881	A-C	5	full-length
AlONSEN10	5	4327023	4327410	A-B	17	solo LTR
AlONSEN11	5	9883484	9888427	A-B	7	full-length
AlONSEN12	6	22653438	22658398	D-A-C	4	full-length
AlONSEN13	7	23781774	23786710	B	7	full-length
AlONSEN14	8	15966326	15970199	A-B		full-length
AlONSEN15	247	3711	8654	A-B	8	full-length
AlONSEN16	1007	2411	3819	A-B	-	incomplete sequenced
AlONSEN17	638	3577	4818	A	12	incomplete sequenced
AlONSEN18	7	8283764	8284127	NA	-	truncated TE
AlONSEN19	1007	42	1687	NA	-	truncated TE
AlONSEN2	1	24919206	24919585	A-C	-	solo LTR
AlONSEN3	2	1120762	1121196	A-C	-	solo LTR
AlONSEN4	2	4268383	4273330	A-C	3	full-length
AlONSEN5	2	12788861	12789817	A-C	-	solo LTR
AlONSEN6	3	13033504	13038475	A-B	2	full-length
AlONSEN7	3	14350695	14355704	A-B	1	full-length
AlONSEN8	3	23055604	23059108	B	-	incomplete sequenced
AlONSEN9	4	16304811	16309766	A-B	10	full-length
AlONSEN	3	14003080	14004319	single palindrome	-	truncated TE
AlONSEN	3	14016614	14019627	single palindrome	-	truncated TE
AlONSEN	1	10332397	10332510	-	-	truncated TE
AlONSEN	1	13615947	13616317	single palindrome	-	truncated TE
AlONSEN	1	24919900	24920073	-	-	truncated TE
AlONSEN	1	25691749	25692168	-	-	truncated TE
AlONSEN	1	25692678	25692869	-	-	truncated TE
AlONSEN	2	11661689	11661997	-	-	truncated TE
AlONSEN	2	10259067	10259437	single palindrome	-	truncated TE
AlONSEN	2	10974958	10975001	-	-	truncated TE
AlONSEN	2	11911062	11911137	single palindrome	-	truncated TE
AlONSEN	3	9491471	9491721	-	-	truncated TE

Name/Family	Scaff.	Start	End	HREs	Locus ID Ito et al. (2013)	Notes
AlONSEN	3	14002611	14002810	A-B	-	truncated TE
AlONSEN	3	23040531	23040611	-	-	truncated TE
AlONSEN	4	15133412	15133767	-	-	truncated TE
AlONSEN	4	15138010	15138365	-	-	truncated TE
AlONSEN	5	18806876	18807231	-	-	truncated TE
AlONSEN	5	9649982	9650265	-	-	truncated TE
AlONSEN	5	9943036	9943399	-	-	truncated TE
AlONSEN	5	11358443	11358638	-	-	truncated TE
AlONSEN	5	12394243	12394562	-	-	truncated TE
AlONSEN	5	12398029	12398350	-	-	truncated TE
AlONSEN	5	12560877	12561493	-	-	truncated TE
AlONSEN	5	13036269	13036319	-	-	truncated TE
AlONSEN	5	18802278	18802633	-	-	truncated TE
AlONSEN	7	3003523	3003877	-	-	truncated TE
AlONSEN	7	8288370	8288733	-	-	truncated TE
AlONSEN	7	14850284	14850720	single palindrome	16	truncated TE
AlONSEN	7	10045372	10046134	single palindrome	-	truncated TE
AlONSEN	7	13857185	13857510	-	-	truncated TE
AlONSEN	7	20135759	20135891	-	-	truncated TE
AlONSEN	7	23674987	23675327	single palindrome	-	truncated TE
AlONSEN	8	6163632	6163987	-	-	truncated TE
AlONSEN	8	6159031	6159386	-	-	truncated TE
AlONSEN	8	15685300	15685663	-	-	truncated TE
AlONSEN	8	17245916	17246222	-	-	truncated TE

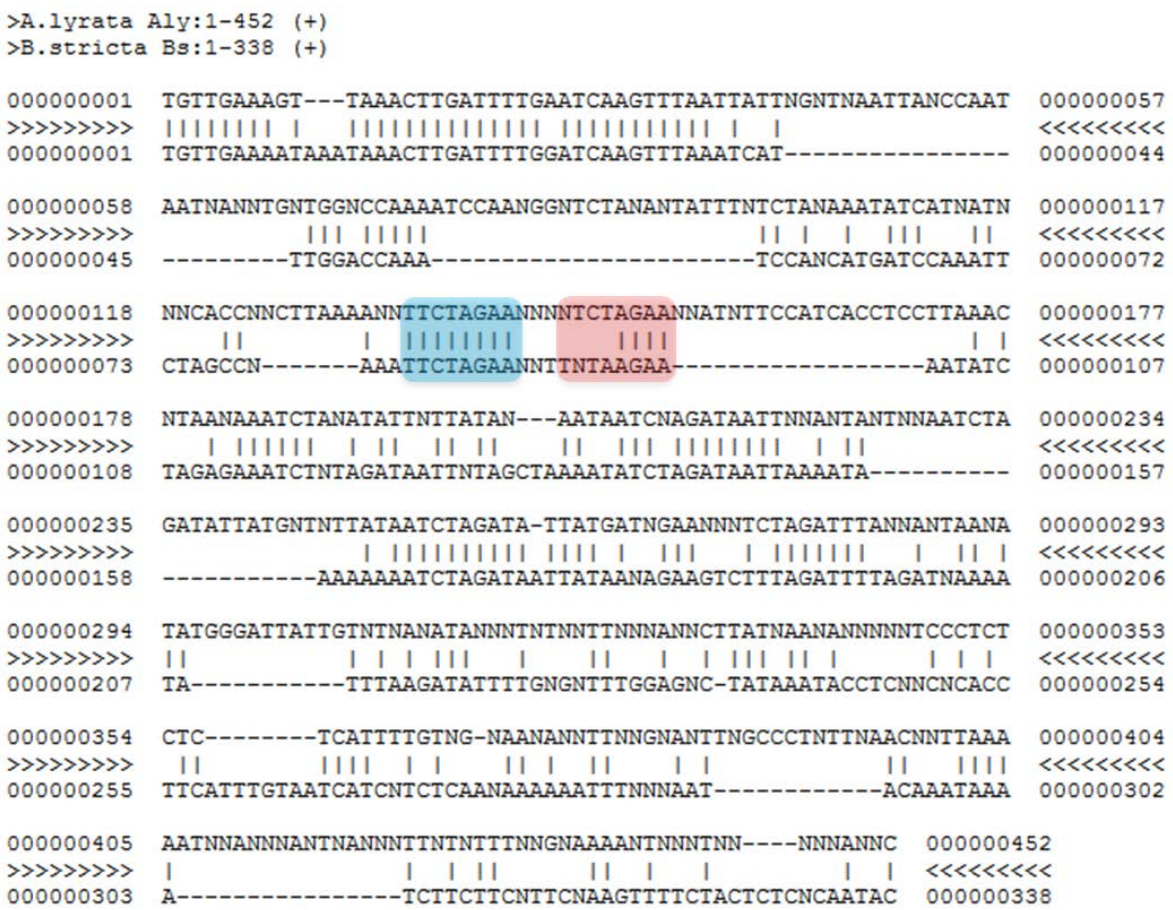

Figure A1: AVID alignment of the *ONSEN* LTR from *Boechera stricta* versus the *ONSEN* LTR of *A. lyrata*. First (blue) and second (red) palindrome is indicated by colors.

Table A6: List of identified *COPIA37* and *HATE* TEs in *Brassicaceae*. Putative HREs are assigned as to their corresponding homolog within *A. lyrata* or novel putative in particular species as X.

Name	Chrom/ Scf	Start	End	Type (solo/full)	LTR size (bp)	LTR identity (%)	HRE
COPIA37							
Arabidopsis lyrata							
AlCOPIA37-1	scf7	7640274	7644088	full	429	98%	C37_1
AlCOPIA37-2	scf3	24130517	24130953	solo	437	-	C37_1–C37_2-
Arabidopsis thaliana							
AtCOPIA37-1	Chr03	10522074	10522491	solo	418	-	-
AtCOPIA37-2	Chr03	12473548	12473995	solo	448	-	C37_1
AtCOPIA37-3	Chr03	12557555	12558039	solo		-	-
AtCOPIA37-4	Chr03	12513617	12514054	solo	439	-	C37_1
AtCOPIA37-5	Chr03	10400149	10400568	Solo	426	-	-
AtCOPIA37-6	Chr03	11393988	11394417	truncated	432	-	C37_1
AtCOPIA37-7	Chr03	20505028	20505461	Solo	435	-	C37_1
AtCOPIA37-8	Chr03	11388241	11387837	Truncated	405	-	-
AtCOPIA37-9	Chr04	1875295	1875724	solo	430	-	-
AtCOPIA37-10	Chr04	5439928	5440332	solo	420	-	-
Capsella rubella							
CrCOPIA37-1	scf1	17530487	17530921	full	409	-	C37_1
CrCOPIA37-2	scf1	14555691	14556103	solo	409	-	C37_1
Capsella grandiflora							
CgCOPIA37-1	scf9816	17768	18180	solo	409	-	C37_1
Eutrema salsugineum							
EsCOPIA37-1	scf1	11995597	11996037	solo	441	-	C37_1 variant
EsCOPIA37-2	scf15	759894	7570335	solo	442	-	C37_1 variant
HATE							
Arabidopsis lyrata							
AlHATE-1	scf1	25445324	25445865	full	525	99	H1-H2
AlHATE-2	scf4	15133415	15138536	full	524	100	H1-H2
AlHATE-3	scf5	9943039	9948175	full	532	99	H1-H2
AlHATE-4	scf7	8283767	8288904	full	532	99	H2
AlHATE-5	scf8	6159034	6164131	full	524	97	H1-H2
AlHATE-6	scf8	15680837	15685660	full	540	99	H2

Name	Chrom/ Scf	Start	End	Type (solo/full)	LTR size (bp)	LTR identity (%)	HRE
Boechera stricta							
BsHATE-1	scf26833	1058555	1059089	solo	525	-	H1
BsHATE-2	scf10040	319604	320075	solo	471	-	-
BsHATE-3	scf4232	581	1095	solo	515	-	X- H1
BsHATE-4	scf30057	1058559	1062295	full	526	94	X- H1
BsHATE-5	scf10199	144407	149499	full	513	96	X- H1
BsHATE-6	scf13129	2245568	2246076	solo	509	-	X- H1
BsHATE-7	scf3288	24153	29242	full	509	96	X- H1
BsHATE-8	scf29223	870868	871384	solo	517	-	-
BsHATE-9	scf3148	824911	830019	full	513	97	X- H1
BsHATE-10	scf18473	1614326	1619381	full	513	96	X- H1
BsHATE-11	scf8819	608944	613973	full	494	98	X- H1
BsHATE-12	scf7867	789356	789865	solo	510	-	H1
BsHATE-13	scf556	4969780	4974903	full	508	95	X
BsHATE-14	scf26959	2009750	2010261	solo	512	-	X- H1
Brassica rapa							
BrHATE-1	Chr09	16190216	16190759	solo	544	-	-
BrHATE-2	Chr09	43527293	43527762	? (too many N)	470	-	-
Eutrema salsugineum							
EsHATE-1	scf8	3962697	3963214	solo	518	-	-
EsHATE-2	scf14	6093144	6098055	full	587	87	-
EsHATE-3	scf5	104664	106190	solo	515	-	-
EsHATE-4	scf15	7311401	7312237	solo	506	-	-
EsHATE-5	scf9	8732118	8732827	solo	455	-	-
EsHATE-6	scf12	9276255	9276948	solo	487	-	-

Abbreviations

A	Adenine
AGO4	Argonaute 4
AP	Aspartic protease
bp	Base pair
C	Cytosine
cDNA	Complementary Desoxyribonucleic acid
CMT3	Chromomethyltransferase 3
Col-0	Arabidopsis thaliana accession Columbia
Col-0	*Arabidopsis thaliana* accession Columbia
DCL3	Dicer –like 3
DDM1	Decreased in demethylation 1
DGE	Differential gene expression
DIR	*Dictyostelium* intermediate repeat
DNA	Desoxyribonucleic acid
dsRNA	Double stranded RNA
epiRILS	epigenetic recombinant inbred lines
ERV	Endogenous retroviruses
et al	et alii / et aliae
et al	et alii / et aliae
EVD	*EVADÉ* (AtCopia93 retrotransposon)
G	Guanine
gDNA	Genomic DNA
H3K9me2	Demethylation of lysine 9 on histone H3
HATE	Heat active transposable element
HRE	Heat Responsive Element
HSE	Heat shock element
HSF	Heat shock transcriptionfactor
HSFA1 a/b/c/d/e; A2	Heat shock transcriptionfactor A1 a/b/c/d/e; A2
INT	Integrase
JA	Jasmonic acid
KYP	Kryptonite
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
LTRE	Low Temperature Responsive Element
Mbp	Mega base pairs
MCIRE	Medicago Cold Inducible Repetitive Element
MET1	Methyltransferase 1
MN47	Arabidopsis lyrata strain MN47
MYA	Million Years Ago
NGS	Next generation sequencing
OPR3	Oxophytodienoate-reductase 3
ORF	Open reading frame
PBS	Primer binding site
PCR	Polymerase chain reaction
PLE	*Penelope*-like elements
PPT	Polypurine tract
qPCR	Quantitative Polmerase chain reaction

qRT-PCR	Quantitative reverse transcriptase PCR
RdDM	RNA directed DNA methylation
RDR2	RNA-dependent polymerase 2
RH	RNase H
RNA	Ribonukleic acid
RNA	Ribonucleic acid
RNAi	RNA interference
RPKM	Reads per kilobase per million
rT	Reverse Transcriptase
SINE	Short interspersed nuclear element
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
ssRNA	Single stranded RNA
T	Thymine
TE	Transposable element
TIR	Terminal inverted repeat
TSS	Transcriptional start site
VLP	Virus-like particle
wt	wild type (plant)

Acknowledgments

First of all I want to thank Dr. Ales Pecinka for the opportunity to work on this exciting topic in his group, his helpful discussions and his support during these years, as well as being assessor in my thesis committee.
I also want to thank Prof. Dr. Maarten Koornneef for the opportunity to do a doctorate in his department and his interest in this work.
I am grateful to Prof. Dr. Achim Tresch at the University of Cologne for his evaluation of this doctoral thesis and for being part of my thesis committee. Furthermore, I would like to thank Prof. Dr. Kay Hofmann for holding the chair of my thesis committee.

This thesis would have never been possible without the help and support of many people. Here, I want to thank Dr. Ahmed Abdelsamad, Navratan Bagwan, Daniel Hamacher and Catarine Markus for practical support and many helpful discussions. In addition, I am very thankful to Prof. Dr. Juliette DeMeaux (University Cologne), Dr. Thomas Piofczyk (MPIPZ Cologne), Prof. Dr. Eric Schranz (University of Wageningen), Prof. Dr. Scott Woody (University of Wisconsin-Madison), Prof. Dr. Detlef Weigel (MPI Tübingen) and Prof. Dr. K.S. Schumaker (University of Arizona) for sharing seeds of the different *Brassicaceae* species with me, which made it possible to test my *in silico* findings in the laboratory.
Besides, I thank Regina Gentges, Manfred Pohe, Barbara Eilts and Ute Tartler for numerous technical supports in the lab and the greenhouses.

I addition, I want to thank all the members of the AG Pecinka, especially Dr. Andreas Finke and Dr. Thomas Piofczyk for all the help in the lab and the great coffee break discussions about science, projects, methods, ongoing experiments and much more. You made me always feel comfortable.

I am very thankful to Dr. Nora Bujdosu for critical reviewing of my thesis.

I am very thankful to my parents and grandparents for all their support, faith and trust in me, always empowering me to stay focused and fulfill my dreams.

Also, great thanks go to the -4-Horsemen, Björn Blöß and Bastian Elßner-Beyer (BEB) for being like a second family of mine, supporting me and for never losing faith in me as a person and scientist.
I am also very thankful to PKW who stood by my side all time during my PhD and supported me with much appreciation and patience.

Special thanks goes to Doris Adler and Prof. Dr. Margarete Baier, without you it wouldn´t have ever been possible for me to start my journey into science.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie abgesehen von unten angegebenen Teilpublikationen noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Maarten Koornneef betreut worden.

Köln, September 2015

Björn Pietzenuk

Lebenslauf

Björn Pietzenuk
Weyerstrasse 11
41363 Jüchen, Deutschland
pietzenuk@mpipz.mpg.de

Persönliche Daten

Geburtstag:	09.09.1982
Geburtsort:	Duisburg (Deutschland)
Staatsangehörigkeit:	deutsche

Schulische Ausbildung

1992 - 2002	Erich Kästner Gesamtschule Homberg, Duisburg Abschluss: Allgemeine Hochschulreife

Akademische Ausbildung

Nov 2011 – Okt 2015	Promotion in der Abteilung für Pflanzenzüchtungsforschung und Genetik am Max-Planck-Institut für Pflanzenzüchtungsforschung, Köln Titel der Doktorarbeit: „Repeated evolution of heat responsivness among *Brassicaceae COPIA* transposable elements”
Okt 2003 - Juli 2011	Studium Diplom Biologie an der Heinrich-Heine-Universität in Düsseldorf am Lehrstuhl für Botanik I Abschlussnote: Diplom (equiv. MSc), 2.27 Titel der Diplomarbeit: “Natural variation and its constraints: Comparison of chloroplast ascorbate peroxidases in higher plants”

Lehre

Okt 2005 – Feb 2010	studentische Hilfskraft im Praktikum “Funktionelle Anatomie und Morphologie der Pflanzen”
Okt 2008 – Feb 2009	Tutor in “Funktionelle Anatomie und Morphology der Pflanzen (Zytologie, Histologie und Organe der Kormophyta)” und “Systematik der Botanik”
April 2008 – Aug 2008	studentische Hilfskraft im Praktikum “Modul Bio 7: DNA fingerprinting, shotgun cloning und Plasmidrestriktion
April 2010 – Aug 2010	studentische Hilfskraft im Praktikum “Modul Bio 3: Zoologie – Stämme des Tierreichs”

Auszeichnungen und Stipendien

Nov 2014 – Okt 2015	Promotionsstipendium der Max-Planck-Gesellschaft