Evolution of the *ONSEN* retrotransposon family activated upon heat stress in Brassicaceae

Hidetaka Ito, Takanori Yoshida, Sayuri Tsukahara, Akira Kawabe

PII: S0378-1119(13)00075-9
DOI: doi: 10.1016/j.gene.2013.01.034
Reference: GENE 38292

To appear in: *Gene*

Accepted date: 17 January 2013

Please cite this article as: Ito, Hidetaka, Yoshida, Takanori, Tsukahara, Sayuri, Kawabe, Akira, Evolution of the *ONSEN* retrotransposon family activated upon heat stress in Brassicaceae, *Gene* (2013), doi: 10.1016/j.gene.2013.01.034

Evolution of the *ONSEN* retrotransposon family activated upon heat stress in Brassicaceae

Hidetaka Ito[1,2], Takanori Yoshida[3], Sayuri Tsukahara[4], and Akira Kawabe[3]

[1] Faculty of Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan

[2] Japan Science and Technology Agency, PRESTO, Kawaguchi, Saitama 332-0012, Japan

[3] Faculty of Life Sciences, Kyoto Sangyo University, Kyoto, Kyoto 603-8555, Japan

[4] Department of Integrated Genetics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

Running title:

Heat-activated retrotransposon in Brassicaceae

Keywords:

Retrotransposon, Brassicaceae, heat stress, evolution

Corresponding author:

Akira Kawabe

Faculty of Life Sciences, Kyoto Sangyo University, Kamigamo Motoyama, Kita-ku, Kyoto, Kyoto 603-8555, Japan

Tel: +81-75-705-3126

Fax: +81-75-705-1914

E-mail: akiraka@cc.kyoto-su.ac.jp

ABSTRACT

A *Ty1*/*Copia*-like retrotransposon, *ONSEN*, is activated subjected to heat stress in *A. thaliana*, and its *de novo* integrations observed preferentially within genes implies its regulation of neighboring genes. Here we show that *ONSEN* related copies were found in most species of Brassicaceae, forming a cluster with each species in phylogenetic tree. Most copies were localized close to genes in *Arabidopsis lyrata* and *Brassica rapa*, suggesting conserved integration specificity of *ONSEN* family into genic or open chromatin. In addition, we found heat-induced transcriptional activation of *ONSEN* family in several species of Brassicaceae. These results suggest that *ONSEN* has conserved transcriptional activation promoted by environmental heat stress in some Brassicaceae species.

1. Introduction

LTR retrotransposons are a major component of plant non-genic genomic regions, and differences in copy number contribute to differences in genome size among species (Hawkins et al., 2006; Vitte and Bennetzen, 2006). The activity of transposons is suppressed epigenetically by DNA methylation and histone modification (Henderson and Jacobsen, 2007; Lisch, 2009; Slotkin and Martienssen, 2007). Despite the tight regulation of transposons in the host genome, most eukaryotes harbor a large number of transposons (Feschotte et al., 2002). Focusing on environmental stress as a trigger for releasing the activation of transposons, we found heat-stressed activation of a *Ty1*/*Copia*-like retrotransposon named *ONSEN* in *Arabidopsis thaliana* (Ito et al., 2011). Activated *ONSEN* not only transcribes but also transgenerationally transposes in heat-stressed plants deficient in siRNAs. Interestingly, in *A. thaliana ONSEN* preferentially inserts within genes (Ito et al., 2011). The heat activation and insertion within or close to genes may confer heat responsiveness to the flanking genes. Indeed, a gene close to new *ONSEN* integration sites responded to heat stress (Ito et al., 2011), suggesting that *ONSEN* insertion can trigger gene network modification.

Changes in gene expression patterns by heat-activated transposons could contribute to adaptation to particular environments. Comparisons of *ONSEN* transposon family across species are thus relevant to understanding evolutionary adaptation to thermal environmental stress. However, all the data of previous experiments was based on only one species, *A.thaliana*. To understand the evolutionary history of the *ONSEN* family and transposon-mediated adaptation, it is necessary to know the distribution of *ONSEN* family and heat activation ability in other species related to *A. thaliana*. In this study, we analyzed the organization and heat-induced activation of *ONSEN* transposon family among several species

of Brassicaceae. Although most species of Brassicaceae have *ONSEN* sequences, some have completely lost *ONSEN* during evolution.

2. Materials and Methods

2.1. Plant materials

In total, 12 species were analyzed: *Arabidopsis lyrata* ssp. *lyrata* (ABRC CS22696, MN47); *A. halleri* ssp. *gemmifera* (Minou, Osaka Japan); *Crucihimalaya wallichii* (SJS00500); *Arabidopsis korshinsky* (SJS00400, syn. *Olimarabidopsis cabulica*); *Turritis glabra* (Ohmi-Shirahama, Shiga, Japan); *Arabis hirsuta* (Kifune, Kyoto Japan); *Lepidium sativum* (purchased from nursery); *Sisymbrium irio* (ABRC CS22563); *Thlaspi arvense* (ABRC CS9661); *Raphanus sativus* (variety Comet); *Capsella rubella* (ABRC CS22697, MTE); and *Cardamine patensis* (purchased from nursery). All plants were grown from seed under a photoperiod of 16 hr light: 8 hr dark. Total DNA was isolated from leaves with the QIAGEN Plant Mini Kit following the manufacturer's instructions.

2.2. Heat treatment and molecular analyses

Plants were grown on MS (Murashige and Skoog) plates with continuous light at 21°C. Seven-day seedlings were subjected to either 21°C (control) or 37°C (heat stress) for 24 hours. Whole plants were sampled immediately after heat treatment and RNA was isolated with TRI Reagent (Sigma). For RT-PCR, the OneStep RT-PCR Kit (QIAGEN) was used with primers Copia1-F (5'-TCTCTG TCT AAT CTA TCA AG-3') and Copia1-R (5'-AAG TTG CTC TAT TGT CAT AGC-3') and the following reaction conditions: 30 min at 50°C; 15 min

at 95°C; 40 cycles of 94°C (30 s), 50°C (30 s), and 72°C (1 min); and 7 min at 72°C. The PCR amplicon size was 363 base pairs.

2.3. Sequence analyses

BLAST searches of the *A. lyrata*, *Brassica rapa*, and *Thellungiella parvula* genomes in GenBank were conducted with *A. thaliana ONSEN* (At1g11265) as the query. To obtain all copies, *A. lyrata* shotgun reads were also searched by BLAST with default search parameters (NCBI Trace Archive Nucleotide BLAST homepage) as described in Tsukahara et al. (2012). When sequences with nucleotide variation that could not be found in assembled genome sequence, shotgun read sequences with the variation were obtained by manual assembly. The genome location was determined by searching assembled genomes by Phytozome ver 8.0 (Goodstein et al. 2012). Because *ONSEN* insertion sometimes caused ambiguous gene annotation in *A. lyrata* genome, we used 700bp flanking sequences in both 5' and 3' of each copy to search in *A. thaliana* genome information (TAIR10) for identification of closest genes.

Part of the reverse-transcriptase (RT) region was amplified from other species of Brassicaceae by PCR (forward primer, CTATTTGCGGAkTGyGAACCyATG; reverse primer, TTGAAATGAGTTGTTGTTGGATGC) and sequenced after cloning; at least 12 clones were sequenced for each species. Identical sequences were considered as the same locus. Newly determined sequences have been deposited in GenBank under accession numbers JX035967-JX035982 and JX975107-JX975185.

2.4. Phylogenetic analyses

Entire *ONSEN* coding sequences from *A. thaliana* and *A. lyrata* were aligned and used for molecular evolutionary analyses with PAML software (Yang, 2007). In this analysis, sequences with long deletion (more than 500bp) were excluded. After exclusion of deletion copies, we used eight *A. thaliana* and 10 *A. lyrata* copies with one *Brassica rapa* sequence as outgroup that is designated as locus1 in Table3. All these sequences include full length coding sequences. The program codeml in PAML with a free-ratio model was used to estimate d_N and d_S values for each branch. The d_N/d_S ratios were compared between species, among internal branches within species, and among external branches within species. The mutations assigned on the internal branches include changes before transposition indicating evolutionary history during active transposable copies, whereas mutation in the external branches is mixture of changes on active and inactive copies. A neighbor-joining tree was constructed for partial *ONSEN* RT region sequences (about 800 bp) from species of Brassicaceae by using synonymous distances (d_S) estimated by the Nei and Gojobori method in MEGA 5 (Tamura et al., 2011).

3. Results and Discussion

3.1. Organization of ONSEN *in* Arabidopsis lyrata

Arabidopsis lyrata is one of the closest relatives of *A. thaliana*, whose genomic sequence has been determined (Hu et al., 2011). Compared with *A. thaliana, A. lyrata* has more transposable elements, possibly due to differences in the mating system (Hollister et al., 2009; Lockton and Gaut, 2010; Tenaillon et al., 2010). In selfing species like *A. thaliana*, insertions of transposable elements into functional genic regions can have severe effects because they will tend to appear in the homozygous condition, whereas in outcrossing species

like *A. lyrata*, even if transposons disrupt function upon insertion into genes, function is retained in the heterozygous condition and the individuals with disrupted genes escape rapid elimination (Charlesworth and Charlesworth, 1995; Wright and Schoen, 1999; Morgan, 2001). Compared with *A. thaliana*, which contains eight full-length copies of *ONSEN*, there are at least 15 *ONSEN*-related sequences in the *A. lyrata* genome; eight are full length and seven are truncated. In addition, there are two solo-LTRs (Table 1). In *A. thaliana*, full-length *ONSEN* copies show no obvious insertion site preference (Ito et al., 2011), although new insertion sites of activated *ONSENs* were located within genic regions in all 11 insertions detected by Ito et al. (2011). The bias in insertion location suggests preferential integration into active chromatin. In *A. lyrata*, *ONSEN* copies are also strongly biased toward insertion into genic regions (Table 1). The proportion of integration into genic regions is high compared to that expected given random insertion, suggesting that *ONSEN* preferentially integrates into genes or open chromatin structures in *Arabidopsis* species.

3.2. Frequent transposition and elimination during evolution

Phylogenetic relationships were estimated based on *Arabidopsis* complete CDS *ONSEN* sequences (Figs. 1A, B). Copies from *A. thaliana* and *A. lyrata* form separate clades in the phylogenetic trees, indicating amplification occurred in each species after species splitting. The branches in *A. lyrata* are long compared to those in *A. thaliana*, although the LTR identities (indicate insertion time) are similar between these two species (Table 1 and Ito et al. 2011: average LTR identities are 0.991 and 0.991 for *A. thaliana* and *A. lyrata*, respectively). One possible explanation is the difference of mating system between the two species. The ONSEN copies of *A. lyrata* could escape elimination because of heterozygous condition to transpose continuously. Among the *ONSEN* copies in *A. lyrata*, 11 have internal

stop codons, frame-shift mutations, and/or a large insertion/deletion, causing interruption of the coding sequence (Table 1). The high frequency of these nonfunctional copies suggests that the mobile copies are rapidly inactivated. The divergence pattern also suggests non-functionalization after transposition, because the d_N/d_S ratio is higher in external branches than in internal branches (Table 2). The low d_N in internal branches suggests conserved protein structure is necessary for active transposition during the process of increase in copy number, whereas the high d_N in external branches indicates mutations to active copies occurred frequently. The mutations could be accumulated because of lower selective constraint, causing degradation and elimination of functional elements.

Individual loci could not increase its frequency due to local elimination and transposition. To analyze the frequency of each *ONSEN* locus within a species, we compared the eight full-length *ONSEN* copies found in one accession, *Columbia*, to 96 *A. thaliana* accessions worldwide (Table S1). All loci showed a low frequency of *ONSEN* insertion, and only two loci were detected in more than 30% of the accessions. The copies AT1g11265, AT3g61330, and AT5g13205, which were present only in one or two accessions, have an identical LTR (Ito et al., 2011), suggesting quite recent transposition. The close phylogenetic relationship among the *ONSEN* copies in *A. thaliana* and the low frequency of each locus indicate that most copies cannot increase its frequency after transposition and are to be degraded. By southern blot analyses, we also confirmed variable copy numbers among *A. thaliana* accessions (three to 20 per accession).

3.3. Evolution of the ONSEN family among species of Brassicaceae

In addition to *A. lyrata*, whole genomic sequences of *Brassica rapa* (Wang et al., 2011) and *Thellungiella parvula* (Dassanayake et al., 2011) have been reported, and we also

obtained *ONSEN* sequences from these species (Table 3). In *B. rapa*, there are 14 *ONSEN*-related sequences. In addition to these copies, there are three solo LTRs in *B. rapa*. Most *B. rapa* copies are located close to genes, similarly to *A. lyrata*. However, *T. parvula* has only three *ONSEN*-related sequences, all of which are nonfunctional copies. To analyze the evolutionary history of *ONSEN* family transposons, we amplified partial *ONSEN* sequences (about 800 bp of the RT region) from 10 other species of Brassicaceae. Among the species we investigated, only *Capsella rubella* and *Cardamine pratensis* showed no PCR amplification of *ONSEN*. This was not due to faulty PCRs or mutations in primer sites, because the intensely sequenced *C. rubella* genome also includes no *ONSEN*-related sequences (NCBI Trace Archive Nucleotide BLAST homepage). To confirm the absence of *ONSEN* sequences in *Capsella* and *Cardamine*, we analyzed *Capsella bursa-pastris* and *Cardamine hirsuta* by PCR. No PCR amplification was observed in either species, suggesting loss of the *ONSEN* family in these two genera.

ONSEN sequences from species of Brassicaceae generally formed clades by species (Fig. 1C). Sequences from *L. sativum*, *A. korshinsky*, *T. glabra*, and *C. walichii*, while grouping by species, in each case formed two distantly related clades. *Arabidopsis halleri* ssp. *gemmifera* and *A. lyrata* were exceptional in that *ONSEN* sequences from the two species were closely related and together comprised a single clade; some of the former comprise 'sister sequences' to the latter and cannot make separate clusters, but some are confined to a clade containing only *A. halleri* or *A. lyrata* sequences (Fig. S1). *A. halleri* and *A. lyrata* are recently diverged sister species in which copies of *ONSEN* present in the common ancestor remain in both daughter species whereas species specific amplification also occurred in these recently diverged species. The phylogenetic relationships among clusters reflected species phylogeny estimated from DNA variations of genic regions (e.g. Johnston et al. 2005, Bailey et al. 2006, Franzke et al. 2011), although species phylogenetic relation is still unclear for

several species used in this study. Sequences from the three *Arabidopsis* species (*A. thaliana*, *A. lyrata*, and *A. halleri*) formed a clade, with *Brassica* and *Raphanus* sequences most distant from this clade. The pattern of *ONSEN* sequences indicates successive 'scrap and build' of *ONSEN* family independently within each species. Possibly, most *ONSEN* copies were eliminated from the genome after transposition and relatively few active copies survived to contribute to the next amplification.

3.4. Heat activation of ONSEN *in species of Brassicaceae*

We previously detected *ONSEN* transcripts in *A. thaliana* seedlings subjected to a temperature shift from 24 h at 6°C to 24 h at 37°C (Ito et al., 2011). The conservation of *ONSEN* sequences among species of Brassicaceae prompted us to examine the activation of *ONSEN* in other species subjected to heat stress. First, we tested heat activation in *A. lyrata* and *A. halleri*, the closest relatives of *A. thaliana*. Using RT-PCR, we detected *ONSEN* transcripts in seedlings of these two species subjected to heat stress, but detected no transcriptional activation in the non-stress condition (Fig. 2A). Although Ito et al. (2011) observed a high frequency of *ONSEN* transposition in the progeny of heat-stressed plants in *A. thaliana*, we could not detect transposition in the progeny of nether heat-stressed nor non-stressed parents in *A. lyrata* (Fig. 2B). Transgenerational retrotransposition had been demonstrated only in an *A. thaliana* mutant deficient in siRNA biogenesis, suggesting that siRNA biogenesis is crucial in suppressing the transposition of *ONSEN* in this species. In addition to full-length transcripts, active *ONSEN* synthesizes extrachromosomal DNA in *A. thaliana* (Ito et al., 2011). To define the activation of *ONSEN* in the progeny of heat-stressed *A. lyrata*, we conducted a Southern blot analysis with non-digested DNA (Fig. 2C). No extrachromosomal DNA was detected, indicating that *ONSEN* was silent in the progeny of

heat-stressed plants.

In addition to *A. lyrata* and *A. halleri*, heat activation was observed in three Brassicaceae species including *A. korshinskyi*, *R. sativus*, and *L. sativum* (Fig. 3). Four species showed *ONSEN* transcripts even in the non-stressed condition including *T. arvense*, *T. glabra*, *C. wallichi*, and *A. hirsuta* (Fig. 3). In these species, activation of several *ONSEN* copies might depend on the constitutive activation of inserted region or the *ONSEN* promoter might contain a homeostatic active domain. We need precise analysis to understand the transcriptional regulation of *ONSEN* among those species. Interestingly, however, heat activation was clear even in species distantly related to *A. thaliana*, such as *R. sativus*. The widely observed heat inducible transcriptional activation indicate that the activation of *ONSEN* is independent of positional or species-specific factors in the *A. thaliana* genome, but rather that a heat-sensitive promoter for *ONSEN* is conserved among some species of Brassicaceae.

3.5. Evolutionary consequences of ONSEN *activation in species of Brassicaceae*

The heat-inducible activation common to *A. thaliana* and some other species of Brassicaceae indicates a conserved heat-responsive cis-element in *ONSEN*. Although all the species in this study naturally inhabit temperate to Polar Regions, environmental conditions similar to the treated heat-stress condition could occasionally occur in summer. Although we were unable to detect transposition induced by heat stress, *ONSEN* had amplified in their evolutional history. *ONSEN* transposition might be regulated by some epigenetic factors. *A. thaliana ONSEN* was transposed by releasing an epigenetic regulation independent of siRNA (HI, unpublished results).

The conserved heat activation and preferential insertion near genic regions among species of Brassicaceae suggest that *ONSEN* could cause changes in gene networks after

higher temperatures resulting from environmental changes. This could ultimately benefit to host species by increasing variation and the ability of populations to adapt to environmental change. Copy numbers, however, are maintained fewer than twenty, in contrast to the high transposition rate in artificially activated copies (Ito et al., 2011). In addition, the high d_N/d_S ratio in external branches, indicative of recent evolutionary events, argues that new insertions are often inactivated and lose retrotransposition activity. The observation suggested that the insertion of *ONSEN* into genic regions might generally have a deleterious effect on the host plants by destructing proper gene networks and/or gene function.

Because amino acid sequences were conserved during both intra- and inter-specific diversification processes, only few functional copies that escaped from degradation could contribute to the active transposition. The transposition-degradation cycle might determine the current *ONSEN* organization in species of Brassicaceae. Once transposable copies are completely lost, the degradation process will eliminate *ONSEN* sequences from the genome of a particular species. In fact, few species of Brassicaceae lack *ONSEN*-related sequences (for example, *Capsella* and *Cardamine*). If the copy number of a transposon family is determined by elimination and transposition rates (Doolittle and Sapienza, 1980; Orgel and Crick, 1980), a high amplification rate (indicated by a star-like phylogeny and species-specific clusters) and faster degradation (due to deleterious effects resulting from the interruption of gene regulation) will produce an equilibrium copy number in a species. Further investigations of *ONSEN* sequences across species and heat responses in species lacking *ONSEN* will be necessary to understand in detail how a retrotransposon family evolves and affects host adaptation.

Acknowledgements

We thank Matthew H. Dick for proofreading the manuscript and Tetsuji Kakutani for assistances. We also thank anonymous reviewers for helpful suggestions. This work was supported in part by a grant from the National Institute of Genetics Cooperative Research Program (2012-B2), a Grant-in-Aid for Young Scientists (B) (23770034 to H.I.), JST, PRESTO to H.I., Grants-in-Aid for Scientific Research in Innovative Areas (23125513 and 23113003 to A.K.; 23119501 to H.I.), and a grant from the Akiyama Life Science Foundation to H.I.

References

Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane, S.L. Jr, Warwick, S.I., Windham, M.D., Al-Shehbaz, I.A., 2006. Toward a global phylogeny of the Brassicaceae. Mol Biol Evol. 23, 2142-2160.

Charlesworth, D., Charlesworth, B., 1995. Transposable elements in inbreeding and outbreeding populations. Genetics 140, 415-417.

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

Dassanayake, M., Oh, D.H., Haas, J.S., Hernandez, A., Hong, H., Ali, S., Yun, D.J., Bressan, R.A., Zhu, J.K., Bohnert, H.J., Cheeseman, J.M., 2011. The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet. 43, 913-918.

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

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

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40, D1178-86.

Henderson, I.R., Jacobsen, S.E., 2007. Epigenetic inheritance in plants. Nature 447, 418-424.

Hollister, J.D., Gaut, B.S., 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 19, 1419-1428.

Hu, T.T., et al., 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat. Genet. 43, 476-481.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., Paszkowski, J., 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472, 115-119.

Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J, Hodnett, G., Drabek, J., Lopez, R., Price, HJ., 2005. Evolution of genome size in Brassicaceae. Ann Bot. 95, 229-35.

Lisch, D., 2009. Epigenetic regulation of transposable elements in plants. Annu. Rev. Plant Biol. 60, 43-66.

Morgan, M.T., 2001. Transposable element number in mixed mating populations. Genet. Res. 77, 261-275.

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

Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8, 272-285.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

Tenaillon, M.I., Hollister, J.D., Gaut, B.S., 2010. A triptych of the evolution of plant

transposable elements. Trends Plant Sci. 15, 471-478.

Tsukahara, S., Kawabe, A., Kobayashi, A., Ito, T., Aizu, T., Shin-I, T., Toyoda, A., Fujiyama, A., Tarutani, Y., Kakutani, T., 2012. Centromere-targeted de novo integrations of an LTR retrotransposon of Arabidopsis lyrata. Genes Dev. 26, 705-713.

Vitte, C., Bennetzen, J.L., 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc. Natl. Acad. Sci. USA 103, 17638-17643.

Wang, X., et al., 2011. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet. 43,1035-1039.

Wright, S. I., Schoen, D.J., 1999. Transposon dynamics and the breeding system. Genetica 107, 139-148.

Yang, Z., 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586-1591.

FIGURE LEGENDS

Fig. 1. Phylogenetic trees for *ONSEN*-related TEs. **A)** d_S tree and **B)** d_N tree based on complete coding sequences, estimated with PAML ver. 4.3 (Yang 2007). The used sequences are labeled as Table1 (eight *A. thaliana* and 10 *A. lyrata*). A *B. rapa* sequence was used for outgroup that is designated as locus 1 in Table3. Bootstrap probabilities more than 75% are shown near the branches. Both trees are shown in a same scale. The scale bar is shown at the bottom left of trees. **C)** d_S tree based on RT region sequences, estimated with MEGA5 (Tamura *et al*. 2011). Bootstrap probabilities of major branches are shown near the branches. The scale bar is shown at the bottom left of tree. Open circles indicate copies in which frame-shift and/or stop codons were found in the analyzed region.

Fig. 2. Effect of heat stress on *ONSEN* activation. **A)** Heat-induced *ONSEN* transcription. Levels of *ONSEN* transcripts in *A. thaliana* (At), *A. lyrata* (Al), and *A. halleri* seedlings (Ah) subjected to 22°C or 37°C were quantified by RT-PCR, with ACTIN2 (ACT2) transcripts used as an internal control. **B)** Southern blot showing the number of *ONSEN* copies in the progeny of heat-stressed plants in *A. lyrata*, indicating that the *ONSEN* copy number did not change in heat-stressed progeny. NS, non-stressed progeny; HS, heat-stressed progeny. B) Genomic DNA of three individual offspring digested with EcoRV. **C)** Southern blot of non-digested DNA loaded to detect extrachromosomal DNA.

Fig. 3. Heat activation of *ONSEN* in species of Brassicaceae. Levels of *ONSEN* transcripts in plants subjected to the control shift or heat stress (+S) were quantified by RT–PCR; 18S ribosomal RNA gene transcripts (18S) were amplified as an internal control.

Fig.1

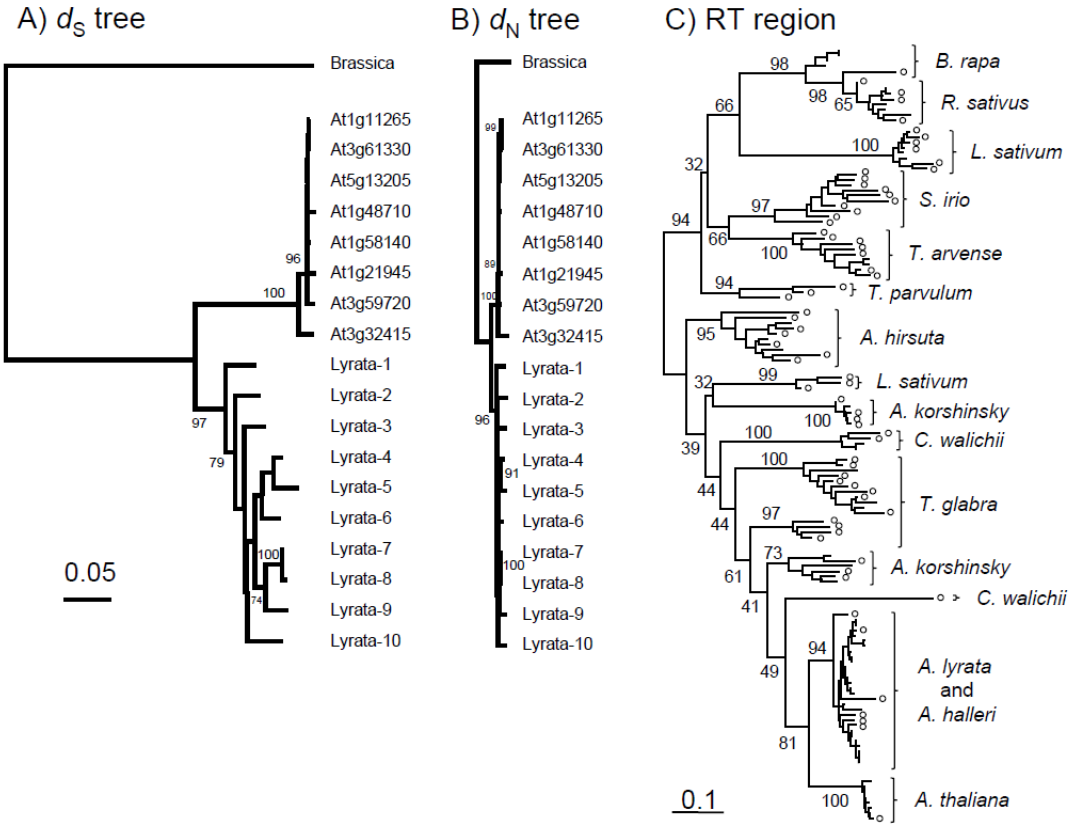

Fig.2

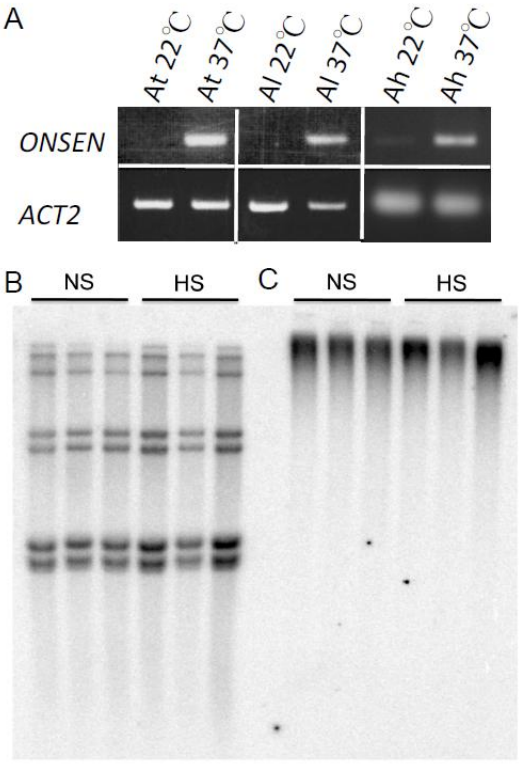

Fig.3

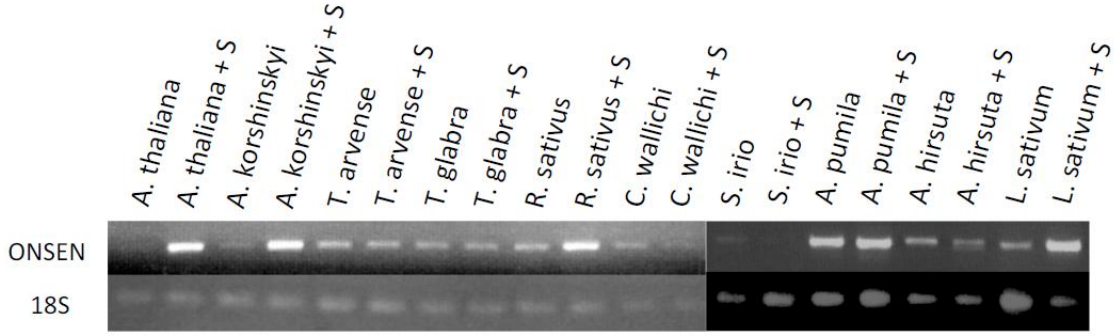

Table 1. List of *A. lyrata ONSEN* related sequences

Locus	Location in *A. lyrata* genome			Closest gene name		Notes	LTR identity
	Scaffold	from	to	ID	Distance		
1	3	14355704	14350695	At4g02050	within	Stop codons	0.982 (442bp)
2	3	13038475	13033504	At3g25490	within		0.991 (448bp)
3	2	4268383	4273330	At1g58340	3975bp 3f	Stop codons	0.988 (424bp)
4	6	22653438	22658398	At2g33480	within		0.991 (444bp)
5	1	11273881	11269188	At1g26510	within	3F LTR truncated	0.976 (125bp)
6	7	23781774	23786710	not found		Stop codon	0.993 (433bp)
7	5	9883484	9888427	At5g66310	within		0.998 (443bp)
8	247	3711	8654	At3g44230	2315bp 3f		0.998 (443bp)
9	Not found			At2g37430	3561bp 3f	3F LTR truncated, Frameshift	0.998 (427bp)
10	4	16304811	16309769	At2g34370	within	Stop codons	0.986 (442bp)
11	Not found			At1g51370	within	Large deletion	0.995 (394bp)
12	638	3577	4819	At1g11145	63bp 5f	3F truncated	-
13	Not found			At1g67870	133bp 5f	5F truncated	-
14	Not found			At5g41880	89bp 3f	5F truncated	-
15	Not found			At5g50080	1395bp 3f	5F truncated	-
16	7	14850720	14850284	At4g07810	42bp 3f	Solo LTR	-
17	5	4327023	4327410	At3g28130	216bp 3f	Solo LTR	-

Some loci were not found in assembled genome (shown as "Not found"). Closest gene that is found in 700bp flanking regions is indicated by TAIR gene ID. Compared nucleotide lengths of LTRs were shown in the parentheses

Table2. Summary of synonymous and nonsynonymous distances

Lineage	Category	Fixed	Internal	External
A. thaliana	dN	0.0072	0.0034	0.0243
	dS	0.1096	0.0101	0.0479
	dN/dS	0.0657	0.3349	0.5081
A. lyrata	dN	0.0055	0.0070	0.0638
	dS	0.0319	0.0796	0.2054
	dN/dS	0.1724	0.0879	0.3107

Distances in every branches were summed to calculate internal and external branch lengths.

Table 3. List of *B. rapa* and *T. parvula ONSEN* related sequences

Locus	Location in the genome			Gene found close to insertion site		Notes	LTR identity
	chromosome	from	to	ID	Distance		
B. rapa							
1	Chr1_848	37464	32659	At1g66610	42bp 3f		0.997 (375bp)
2	Chr4_251	22493	17939	At4g11000	within		0.997 (399bp)
3	Scaffold000140_9	6959	2047	At5g61700	6bp 5f		0.975 (405bp)
4	Chr9_299	5771	10353	At2g33200	intron		0.992 (398bp)
5	Chr9_1312	90899	95273	At3g61730	250bp 3f		0.958 (377bp)
6	Chr9_445	66309	63812	At4g10510	700bp 3f	deletion	0.997 (399bp)
7	Chr9_445	68823	66459	At4g10110	400bp 3f	share LTR with locus 6	-
8	Chr10_432	92425	89277	-		deletion	0.871 (31bp)
9	Chr6_164	19600	21895	At1g17480	600bp 5f	5F truncated	-
10	Chr4_51	21537	19258	-		deletion	0.971 (376bp)
11	Chr5_824	63702	61082	At3g17230	within	deletion	0.995 (398bp)
12	Chr8_395	36218	39715	At1g31880	150bp 3f	large insertion, deletion	0.912 (387bp)
13	Chr4_370	41198	45902	At1g04170	200bp 5f	5F truncated	1.000 (398bp)
14	Chr5_317	26868	27984	-		5F truncated	-
15	Chr3_517	5501	5143	-		solo LTR	-
16	Chr2_52	21253	21654	-		solo LTR	-
17	Chr2_690	4139	4505	At2g03290	100bp 3f	solo LTR	-
T. parvula							
1	chr2-2-c15	1181462	1186770	-		5F truncated	0.964 (223bp)
2	chr1-5-c8	5196445	5191607	-		5F truncated, deletion	0.860 (207bp)
3	chr7-4-c6	2731950	2711343	-		large insertion, deletion	0.891 (411bp)

Closest gene that is found in 700bp flanking regions is indicated by TAIR gene ID. Compared nucleotide lengths of LTRs were shown in the parentheses

Abbreviations

LTR, long terminal repeat; ABRC, Arabidopsis biological resource center; BLAST, basic local alignment search tool; TAIR, the Arabidopsis information resource; RT, reverse transcriptase; CDS, coding DNA sequence; TE, transposable element

Highlights

- We analyzed evolution of a retrotransposon *ONSEN* family in Brassicaceae.
- Most species of Brassicaceae have *ONSEN* family.
- The *ONSEN* family sequences locate within or close to genes.
- Heat-induced activation was observed widely in Brassicaceae.