ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2014) **12**(S1); S146–S150
doi:10.1017/S1479262114000446

Comparative analysis of *Cassandra* TRIMs in three Brassicaceae genomes

Perumal Sampath[1] and Tae-Jin Yang[1]*

[1]*Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea*

Abstract

Terminal-repeat retrotransposon in miniature (TRIM) elements are a miniature form of retrotransposons and play an important role in genome organization. The *Cassandra* TRIM family has been identified in over 50 plant species, including both monocots and dicots. *Cassandra* elements carry an independently transcribed 5S RNA sequence in their terminal repeat regions, which is unique compared with other TRIM families. Although the existence of *Cassandra* elements has been documented in many plants, much work remains to characterize *Cassandra* family members and elucidate their distribution. In this study, we comparatively analysed the *Cassandra* family members in the *Brassica oleracea*, *B. rapa* and *Arabidopsis thaliana* genomes. A total of 602, 451 and 173 members, of which 130, 60 and 9 were relatively intact, were identified from the *B. oleracea*, *B. rapa* and *A. thaliana* genomes, respectively. Most of the *Cassandra* elements (1120/1226) were found in intergenic spaces, but 106 elements were inserted in genic regions such as introns, exons and untranslated regions. Our comparative analysis of the *Cassandra* family members in *A. thaliana*, *B. rapa* and *B. oleracea* reveals that some *Cassandra* elements have been commonly retained during the last 20 million years in three species and some elements have been uniquely evolved in *Brassica* species. This study promotes our understanding of the role and utility of *Cassandra* elements in the evolution of the Brassicaceae family.

Keywords: *Arabidopsis thaliana*; *Brassica*; *Brassica oleracea*; *Brassica rapa*; *Cassandra*; TRIMs

Introduction

Terminal-repeat retrotransposon in miniature (TRIM) elements are small (<900 bp) in size, abundant and ubiquitously present in plant genomes, in addition to being found in the genome of an ant, *Pogonomyrmex barbatus* (Witte *et al.*, 2001; Zhou and Cahan, 2012). TRIM elements share structural features with long terminal-repeat (LTR) retrotransposons, including the terminal repeat (TR), primer-binding site (PBS) and polypurine tract (PPT); thus, TRIM elements are considered to be derivatives of LTR retrotransposons (Wicker *et al.*, 2007; Zou *et al.*, 2009). TRIM elements sometimes associate with genic regions and can change gene structure and function by affecting the promoter, shuffling the coding region and/or altering gene expression. In addition, TRIM elements are an important source of molecular markers that have been effectively utilized for genome mapping, diversity and evolutionary studies (Witte *et al.*, 2001; Yang *et al.*, 2007).

The Brassicaceae family is economically important, with >330 genera and around 3700 species. It has diverse phenotypic resources and includes important species for studies of the evolution of polyploidy (Johnston *et al.*, 2005; Mun *et al.*, 2009; Wang *et al.*, 2011). In addition, the widely used diploid model plant, *Arabidopsis thaliana*, is a member of the Brassicaceae family. The *Brassica rapa* and *B. oleracea* genomes have now been sequenced and are publicly available, revealing that around 40% of both genomes is derived

* Corresponding author. E-mail: tjyang@snu.ac.kr

from transposable elements (Cheng *et al.*, 2011; Wang *et al.*, 2011; Yu *et al.*, 2013). Two studies have suggested that TRIM elements may play active roles in the evolution of duplicated genes in the *B. rapa* genome (Kwon *et al.*, 2007; Yang *et al.*, 2007).

Members of a particular TRIM family, *Cassandra*, have been identified in many plant genomes including *Brassica* species. *Cassandra* elements have a unique TR structure in which some carry a highly conserved *5S rRNA* gene. Moreover, *Cassandra* TRs are very similar in structure in monocot and dicot plants (Kalendar *et al.*, 2008). In this study, we comparatively analysed the *Cassandra* family members in the *A. thaliana*, *B. oleracea* and *B. rapa* genomes and revealed their continuous and recent amplification in each *Brassica* species.

Materials and methods

Sequence analysis of Cassandra *elements from Brassicaceae species*

The sequences of *Cassandra* family members from *B. oleracea* (AY860308), *B. rapa* (AY860309) and *A. thaliana* (AY923743) were used as queries for the analysis of *Cassandra* elements in the three genomes (Kalendar *et al.*, 2008). All the *Cassandra* elements were retrieved from the three genomes using BLASTn analysis with an *E* value $< \times 10^{-10}$ from a local database (http://im-crop.snu.ac.kr/). The genome sequences and gene annotation information of *B. oleracea* (version 1.0), *B. rapa* (version 1.2) and *A. thaliana* (version TAIR10) were obtained from Bolbase (Yu *et al.*, 2013), BRAD (Cheng *et al.*, 2011) and TAIR (Lamesch *et al.*, 2012), respectively. The distributions of *Cassandra* elements in the *B. oleracea*, *B. rapa* and *A. thaliana* genomes were determined using an *in silico* mapping tool (Sampath *et al.*, 2013). The insertion positions of *Cassandra* elements in various genomic locations such as introns, exons, untranslated regions and intergenic spaces in *B. oleracea*, *B. rapa* and *A. thaliana* were characterized using a custom Perl script. The nearly intact *Cassandra* family members (defined as having >80% sequence similarity and >80% sequence coverage) from the *B. oleracea*, *B. rapa* and *A. thaliana* genomes were used for phylogenetic analysis, and the phylogenetic tree was constructed using MEGA5 (Tamura *et al.*, 2011).

Results and discussion

We retrieved all the *Cassandra* elements in the *A. thaliana*, *B. oleracea* and *B. rapa* genomes (Kalendar *et al.*, 2008). The retrieved elements included the conserved signature domains, namely the TR

Table 1. Characteristics of *Cassandra* elements in *Arabidopsis thaliana*, *Brassica oleracea* and *B. rapa*

TRIM	Total size (bp)	TR (bp)	TSD (bp)	AT (%)	TRIM members[a]			TRIM member distribution[b]					
					Full	Solo TR	Truncated	Intergenic space	300 bp upstream	CDS	Intron	300 bp downstream	Total
At-*Cassandra*	824	356	5	46	9	20	144	155	1	11	2	4	173
Bo-*Cassandra*	801	349	5	49	130	108	364	567	12	0	13	10	602
Br-*Cassandra*	809	350	5	47	61	89	301	398	5	1	42	5	451

TRIM, Terminal-repeat retrotransposon in miniature; TR, Terminal repeat; TSD, Target Site Duplication; CDS, Coding DNA Sequence.
[a] At-*Cassandra*, Bo-*Cassandra* and Br-*Cassandra* members were extracted from 120, 256 and 385 Mb of pseudo-chromosome sequences from *A. thaliana*, *B. oleracea* and *B. rapa*, respectively. The detailed physical position information and characterization of members are summarized in Tables S1, S2 and S3 (available online) for *A. thaliana*, *B. oleracea* and *B. rapa*, respectively. [b] Genomic distributions of *Cassandra* elements from *A. thaliana*, *B. oleracea* and *B. rapa* were characterized using information from TAIR (The *Arabidopsis* Information Resource; Lamesch *et al.*, 2012), Bolbase (Yu *et al.*, 2013) and BRAD (Cheng *et al.*, 2011), respectively.

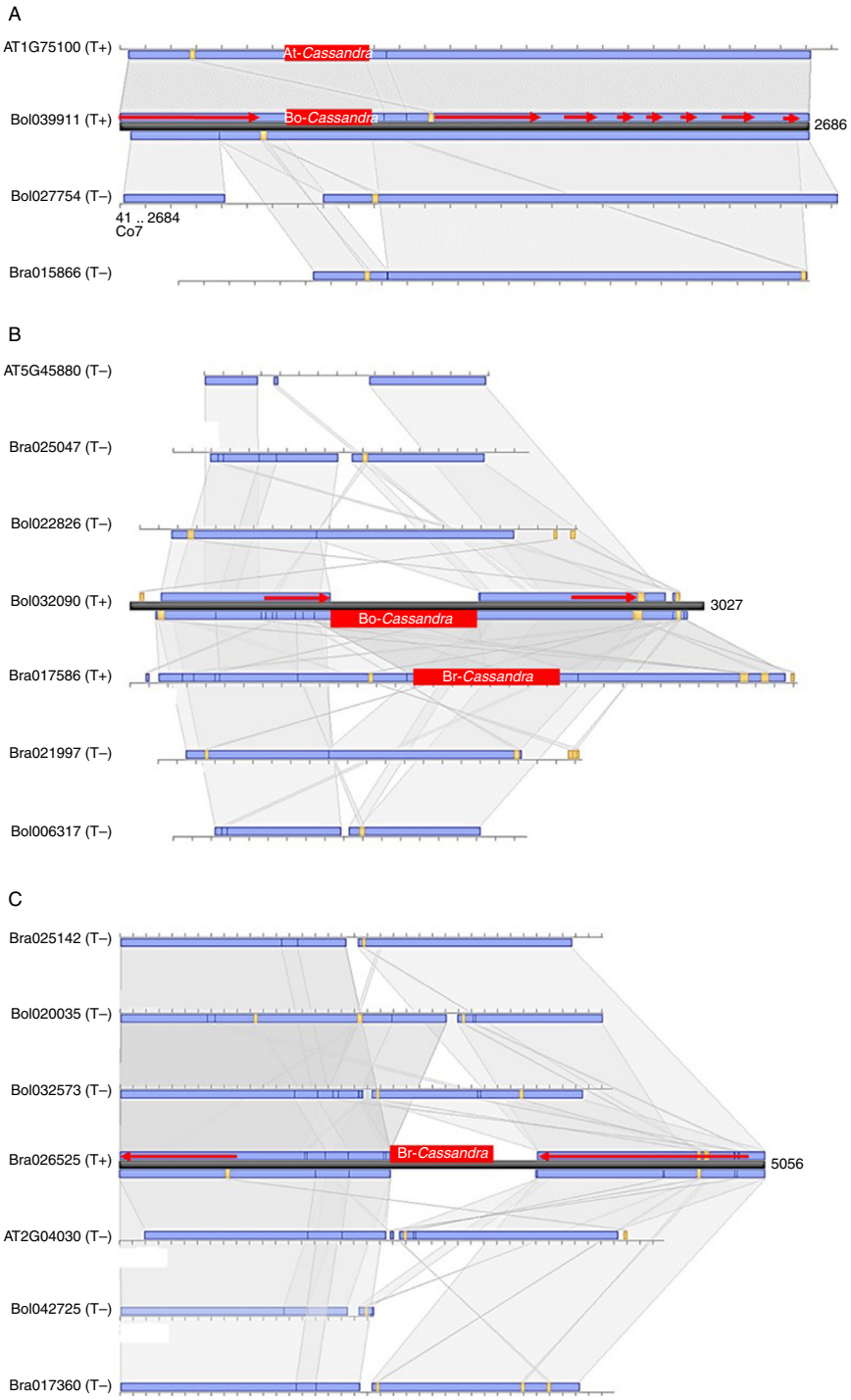

Fig. 1. Micro-synteny comparison of genomic regions containing *Cassandra* elements with their non-inserted paralogues (NIPs) and non-inserted orthologues (NIOs) in *Arabidopsis thaliana, Brassica oleracea* and *B. rapa*. (A) Shared insertion in *A. thaliana* and *B. oleracea*. Micro-synteny comparison of genomic regions of At-*Cassandra* (AT1G75100) and Bo-*Cassandra* (Bol039911) with its NIP (Bol027754) and a NIO from *B. rapa* (Bra015866). (B) Shared insertion in *B. oleracea* and *B. rapa*. Micro-synteny comparison of genomic regions of Bo-*Cassandra* (Bol032090) and Br-*Cassandra* (Bra017586) with its NIPs (Bol006317 and Bol022826) and NIOs from *B. rapa* (Bra025047 and Bra017586) and *A. thaliana* (AT3G45880). (C) Unique *Cassandra* element insertion in *B. rapa*. Micro-synteny comparison of the genomic region of Br-*Cassandra* (Bra026525) with its NIPs (Bra025142 and Bra017360) and NIOs from *B. oleracea* (Bol020035, Bol032573 and Bol042725) and *A. thaliana* (AT2G04030). Exons and gene direction are indicated with red arrows. *Cassandra* element insertions are shown as red bars. T + and T − indicate genes with *Cassandra* insertion and non-insertion, respectively.

(~350 bp), PBS, PPT and internal sequence. The sizes of the *Cassandra* family members ranged from 801 to 826 bp, and the elements had a low AT content (45–49%) compared with the whole genomes (>63.8%; Wang *et al.*, 2011). Hereafter, we will refer to the *Cassandra* TRIM families from *A. thaliana*, *B. oleracea* and *B. rapa* as At-*Cassandra*, Bo-*Cassandra* and Br-*Cassandra*, respectively (Table 1). Most of the Bo-*Cassandra* and Br-*Cassandra* members shared high sequence similarity. By contrast, At-*Cassandra* members exhibited dissimilarities and indel structures compared with Bo-*Cassandra* and Br-*Cassandra* members, suggesting that the *Cassandra* family has recently been amplified in the *Brassica* genus (Fig. S1, available online).

A. thaliana, *B. oleracea* and *B. rapa* pseudo-chromosome sequences contained 173, 602 and 451 members, respectively. Among the 602 Bo-*Cassandra* members, 130, 108 and 364 had relatively intact elements, solo TR and fragmented structures, respectively. Of the 451 Br-*Cassandra* members, 60 were intact. Only 9 of the 173 At-*Cassandra* members were identified as intact, suggesting that *Cassandra* elements proliferated in the *Brassica* genus after divergence from the *Arabidopsis* genus (Table 1; Tables S1, S2 and S3, available online). We also analysed the *Cassandra* family members in *Oryza sativa*, *Medicago truncatula* and *Zea mays* and identified 32, 10 and 1087 intact elements, respectively (Tables S4, S5 and S6, available online).

Cassandra family members, including intact elements, solo TR and fragments of *Cassandra* elements, exhibited random distribution patterns on the *A. thaliana*, *B. oleracea* and *B. rapa* pseudo-chromosome sequences (Fig. S2, available online). In addition, a survey of only intact elements also revealed a random distribution pattern. The characterization of insertion positions revealed that the *Cassandra* family members preferentially resided in intergenic spaces (88–94%), although a number of these were located in the introns of genes in *B. oleracea* and *B. rapa* (Table 1; Tables S1, S2 and S3, available online).

Our phylogenetic analysis revealed that At-*Cassandra* members are diverse compared with Bo-*Cassandra* and Br-*Cassandra* members. The high sequence conservation of Bo-*Cassandra* and Br-*Cassandra* elements suggests that Bo-*Cassandra* and Br-*Cassandra* members have recently been amplified in the *Brassica* genus (Fig. S3, available online).

Comparative analysis of homologous sequences harbouring *Cassandra* members can reveal their transposition period (Kwon *et al.*, 2007; Yang *et al.*, 2007; Sampath *et al.*, 2013). One of the *Cassandra* members was found in the syntenic regions of the three genomes, suggesting that it might have evolved before the *A. thaliana*–*Brassica* split 20 million years ago (MYA; Fig. 1(A); Mun *et al.*, 2009; Town *et al.*, 2006; Yang *et al.*, 1999). Other members were unique in the *Brassica* genus, coexisting in one of the triplicated paralogous regions of *B. oleracea* and *B. rapa*. These elements probably inserted themselves into the region between 4 and 17 MYA, after genome triplication of the tribe Brassicaceae but before speciation of *B. oleracea* and *B. rapa* (Fig. 1(B); Mun *et al.*, 2009; Town *et al.*, 2006; Yang *et al.*, 1999). Another Br-*Cassandra* member was unique to *B. rapa*, suggesting that it was activated after the divergence of *B. rapa* and *B. oleracea* 4 MYA (Fig. 1(C); Mun *et al.*, 2009; Town *et al.*, 2006; Yang *et al.*, 1999). As exemplified by these three distribution patterns, *Cassandra* elements have remained active in these species for more than 20 million years.

The present study reveals the continuous activation of *Cassandra* TRIM family elements throughout evolution in Brassicaceae species, even though the *Cassandra* family was derived from the ancestor of monocot and dicot plants. Further identification of autonomous partner elements will be important for exploring the functional and evolutionary roles of *Cassandra* elements in plants.

Supplementary Material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262114000446

Acknowledgements

This work was supported by Golden Seed Project (Center for Horticultural Seed Development, No. 309008-05-1-cg000), Ministry of Oceans and Fisheries (MOF), Republic of Korea.

References

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W and Wang X (2011) BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biology* 11: 136.

Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R and Price HJ (2005) Evolution of genome size in Brassicaceae. *Annals of Botany* 95: 229–235.

Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O and Schulman AH (2008) Cassandra retrotransposons carry independently transcribed 5S RNA. *Proceedings of the National Academy of Sciences of the United States of America* 105: 5833–5838.

Kwon SJ, Kim DH, Lim MH, Long Y, Meng JL, Lim KB, Kim JA, Kim JS, Jin M, Kim HI, Ahn SN, Wessler SR, Yang TJ and Park BS (2007) Terminal repeat retrotransposon in miniature (TRIM) as DNA markers in *Brassica* relatives. *Molecular Genetics and Genomics* 278: 361–370.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A and Huala E (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202–D1210.

Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I and Park BS (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biology* 10: R111.

Sampath P, Lee S-C, Lee J, Izzah NK, Choi B-S, Jin M, Park B-S and Yang T-J (2013) Characterization of a new high copy Stowaway family MITE, BRAMI-1 in *Brassica* genome. *BMC Plant Biology* 13: 56.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.

Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, Vigouroux M, Trick M and Bancroft I (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *The Plant Cell* 18: 1348–1359.

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Yu J, Meng J, Min J, Poulain J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Z, Li Z, Xiong Z and Zhang Z (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* 43: 1035–1039.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P and Schulman AH (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: 973–982.

Witte C-P, Le QH, Bureau T and Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proceedings of the National Academy of Sciences of the United States of America* 98: 13778–13783.

Yang TJ, Kwon SJ, Choi BS, Kim JS, Jin M, Lim KB, Park JY, Kim JA, Lim MH, Kim HI, Lee HJ, Lim YP, Paterson AH and Park BS (2007) Characterization of terminal-repeat retrotransposon in miniature (TRIM) in *Brassica* relatives. *Theoretical and Applied Genetics* 114: 627–636.

Yang Y-W, Lai K-N, Tai P-Y and Li W-H (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *Journal of Molecular Evolution* 48: 597–604.

Yu J, Zhao M, Wang X, Tong C, Huang S, Tehrim S, Liu Y, Hua W and Liu S (2013) Bolbase: a comprehensive genomics database for *Brassica oleracea*. *BMC Genomics* 14: 664.

Zhou Y and Cahan SH (2012) A novel family of terminal-repeat retrotransposon in miniature (TRIM) in the genome of the red harvester ant, *Pogonomyrmex barbatus*. *PLoS One* 7: e53401.

Zou J, Gong H, Yang T-J and Meng J (2009) Retrotransposons – a major driving force in plant genome evolution and a useful tool for genome analysis. *Journal of Crop Science and Biotechnology* 12: 1–8.