Molecular analysis of a novel DNA transposon in Triticeae

Thiyagarajan K., Cantale C., Galeffi P.

in

Porceddu E. (ed.), Damania A.B. (ed.), Qualset C.O. (ed.).
Proceedings of the International Symposium on Genetics and breeding of durum wheat

Bari : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 110

2014
pages 613-617

Molecular analysis of a novel DNA transposon in *Triticeae*

Karthikeyan Thiyagarajan[1], Cristina Cantale[1], Enrico Porceddu[2], Patrizia Galeffi[1]

[1] ENEA CR Casaccia, UTAGRI-INN, Rome, Italy
[2] University of Tuscia, Viterbo, Italy

Abstract. A novel non-autonomous transposon element was identified in durum wheat and in various *Aegilops speltoides* accessions from the Fertile Crescent. It shows the standard transposon signals, such as a terminal inverted repeat (TIR-18bp), target site duplications (TSD-2bp-TC), many internal inverted repeats and a variable number of tandem repeats and does not code for a transposase enzyme. Interestingly, it is located inside the *Dehydration Responsive Factor 1* (*TdDRF1*) gene which codifies for transcription factors involved in the early response to drought by an alternative splicing mechanism. The transposon encompasses the gene sequence, from intron 1 to intron 3, including two translated regions, the exon 2 and the exon 3. Due to its peculiar position inside the CDS, a possible involvement in the molecular evolution of the gene was hypothesized. The transposon sequence and signals in all available relevant sequences from the same tr be, such as *Triticum durum, T. urartu, A. tauschii* were analysed and compared with the aim of drawing its phylogenetic story.

Keywords. Transposable Elements – DRF1 gene – Molecular evolution – Exonization.

Analyse moléculaire d'un nouveau transposon d'ADN chez les Triticées

Résumé. *Un nouvel élément transposon non autonome a été identifié chez le blé dur et chez diverses accessions d'Aegilops speltoides du Croissant Fertile. Il montre les signaux de transposons standards tels qu'une répétition terminale inversée (TIR-18BP), des duplications du site cible (TSD-2pb-TC), de nombreuses répétitions inversées internes et un nombre variable de répétitions en tandem et il ne code pas pour une enzyme transposase. Il est intéressant de noter que cet élément est situé à l'intérieur du gène Dehydration Responsive Factor 1 (TdDRF1) qui code pour les facteurs de transcription impliqués dans la réponse précoce à la sécheresse par un mécanisme d'épissage alternatif. Le transposon comprend la séquence du gène, de l'intron 1 à l'intron 3, incluant deux régions traduites, l'exon 2 et l'exon 3. En raison de sa position particulière à l'intérieur du CDS, une possible implication dans l'évolution moléculaire du gène a été avancée. La séquence du transposon et les signaux dans toutes les séquences pertinentes disponibles de la même tribu, comme Triticum durum, T. urartu, A. tauschii ont été analysés et comparés en vue de tracer son histoire phylogénétique.*

Mots-clés. *Eléments transposables – Gène DRF1 – Evolution moléculaire – Exonisation.*

I – Introduction

Transposable Elements (TEs) are genetic elements capable of transposing to different chromosomal locations and represent a large portion of the DNA in many species of animals and plants including agriculturally important crops such as corn and wheat (SanMiguel *et al*., 1996). Transposons are classified into two classes according to their mechanism of transposition: Class I – retrotransposons and Class II – DNA transposon. In particular, the DNA transposons are excised from one to another place with simple cut and paste mechanism. Based on the coding ability of transposase, the DNA transposons are categorized as autonomous and non-autonomous. Initially, it was supposed that TEs were simply 'junk' DNA, but subsequently it was demonstrated that they play an important role in evolution and speciation through mechanisms such as exonization and intronization (Fedoroff, 2000; Sorek, 2007; Sela *et al*., 2010; Chenais *et al*., 2012).

We identified a novel non-autonomous transposon element located inside the *Dehydration Responsive Factor 1* (*DRF1*) gene, a *DREB2*-related gene correlated to drought stress response in wheat. The gene consists of four exons and three introns and its expression is modulated by an alternative splicing mechanism (Latini *et al.,* 2007). Homologous genes sharing the same structure were isolated and analysed in wheat wild relatives, *Triticum urartu, Aegilops speltoides and Aegilops tauschii* and also in other *Poaceae* family members, mining sequences databases. The transposon inside *DRF1* gene was analysed for investigating its role in the gene evolution.

II – Material and methods

1. Transposon Mining

Genomic sequences of *DREB2*-related genes were accessed from GenBank (National Center for Biotechnology Information, NCBI; http:www.ncbi.nlm.nih.gov), Phytozome v9.1 (http://www.Phytozome.org) and TAIR (http://www.arabidopsis.org) databases, between January and June 2013. *AsDRF1* transposon sequence and *TdDRF1* gene sequence were used as BLAST search queries (version 2.2, http:blast.ncbi.nlm.nih.gov/Blast.cgi?).

2. Data analysis

Recovered sequences of DREB2-related genes were analysed by CLUSTAL W (Thompson *et al.,* 1994) and functionally aligned by BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Molecular evolutionary genetics analyses were carried out by MEGA 5 (Tamura *et al.,* 2011).

III – Results and discussion

1. The *DRF1* transposon and a hypothetical mechanism for transposon insertion

The analysis of the *DRF1* gene sequence in durum and its ancestors revealed some inverted repeats, target site duplications and the presence of many internal reverse and direct short tandem and long tandem repeats, all signals of transposable elements. In particular, it was observed a terminal inverted repeat (TIR-32bp) anchoring another internal 100% identity TIR-18bp, plus target site duplications (TSD-2bp-TC). Thus, a new transposon was identified and added to Repbase (Karthikeyan *et al.,* 2009). Because no sequence coding for a transposase enzyme was found, it represents a non-autonomous element. This transposon encompasses the *DRF1* gene sequence, from intron 1 to intron 3 and includes two translated regions, the exon 2 and the exon 3, suggesting a transposition event followed by exonization.

The overall structure of the transposon and the alignment of TIRs and TSDs in wheat and its ancestors are shown in Figure 1.

The transposon structure strongly supports that it could have played a vital role in the gene evolution. Actually, *DREB2* gene, firstly isolated in *Arabidopsis thaliana,* do not exhibit such a structure.

Based on the similarity between the 32bp and the 18bp TIRs, a double-step event for transposition can be hypothesized in this group of sequences. The 32bp TIRs suggest an older event of insertion of a non-autonomous transposon about 255bp in an ancestral gene sequence. It is possible to hypothesize that later, a transposase, carrying about 1190bp transposon containing 18bpTIRs and 4bp STRs, due to the high similarity of TIRs, inserted it in the target site. The final assembled gene structure is schematically shown in Figure 2.

Figure 1. Structure and alignment of 18bp TIRs and 32bp TIRs in *DRF1* genes from different *Poaceae* members. TIRs are shown in red colour, TSDs are shown in yellow colour. A.c: acc. GU017675; A.t: acc. EU197052; T.d: acc. EU197052; Td2: acc. JN571425; T.a: sequence from Chr. AL at URGI (http://urgi.versailles.inra.fr); T.u: lab sequence acc. 57_7; As1: acc. FJ843102; As2: acc. FJ858188; As3: acc. FJ858187.

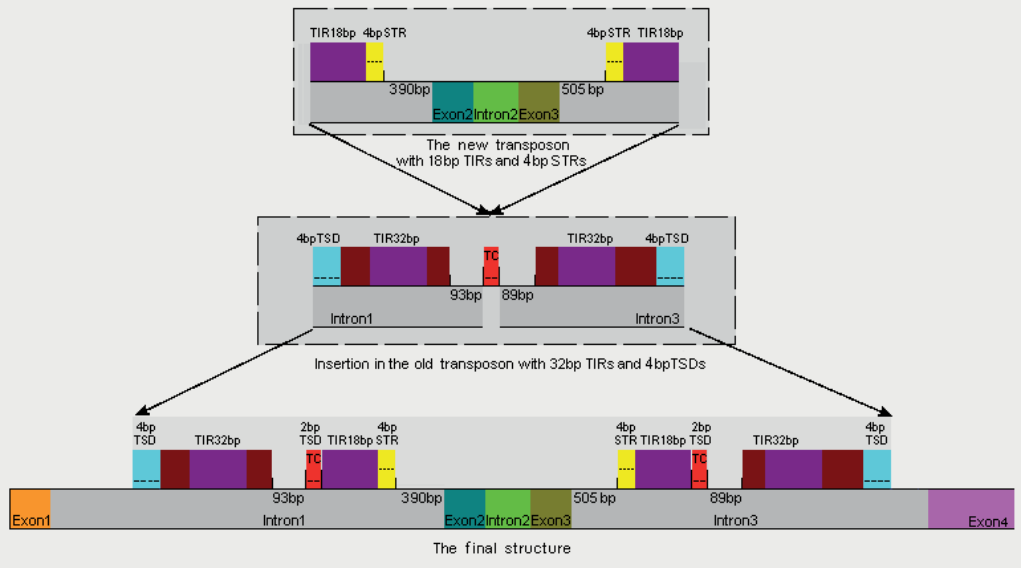

Figure 2. Hypothetical mechanism of the insertion of transposon inside *DRF1* gene sequence (the scheme is based on acc. FJ843102).

3. Analysis of *DRF1* transposon in orthologous sequences

Available databases were searched for retrieving sequences of orthologous DREB2-related genes. Just few genomic DNA sequences were available, being the great majority mRNA sequences. Sequences were functionally analysed and compared and it appears that only sequences from *Poaceae* family showed the same *DRF1* gene structure. In this subset, the sequences of transposon were isolated and aligned by ClustalW. A phylogenetic tree was built from the alignment (see Figure 3).

It is worth of noting that the obtained phylogenetic tree, able to clusterize correctly the species, exactly reflects the plant taxonomy.

Evolutionary analysis was carried out using 844 positions after eliminating gaps and missing data. The probability of transition resulted higher than transversion, thus suggesting a strong selection pressure to promote diversification.

Concerning TIRs, they appear to be well conserved in wheat and its wild relatives, being better conserved at 5', because, as known, the 3' TIRs are more susceptible to mutate. The core element, constituted by exon 2-intron 2-exon 3, is largely conserved in all analysed sequences, probably because it corresponds to the part of transposon which acquired a functional role in the gene. Concerning more distant species, just the 18bp TIRs can be recognized, thus the double transposition event cannot be hypothesized, and only the later insertion can be observed.

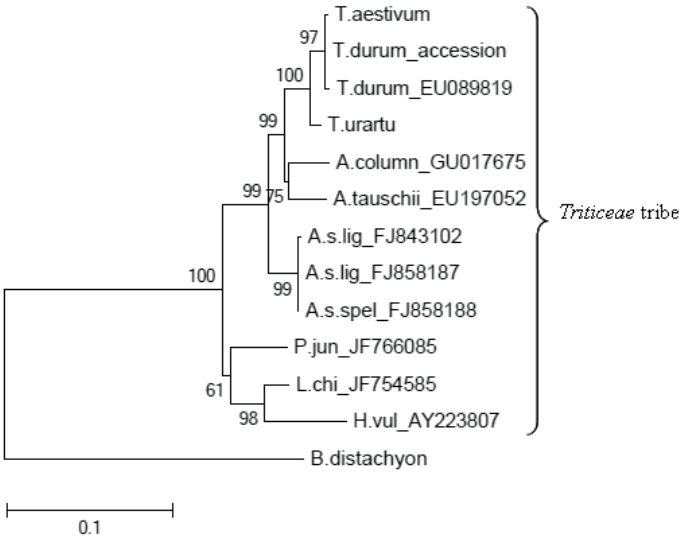

Figure 3. Phylogenetic tree from the alignment of the orthologous sequences of *DRF1* TE from T*riticeae* tribe and *B. distachyon* (Maximum likelihood method, Kimura 2 parameter model, 1000 bootstrap).

4. Looking for the ancestral gene sequence

Looking at the possible mechanism of transposon insertion, we speculated a double-step transposition inside an ancestral sequence. Thus, to mimics this ancestral sequence, we manually removed the whole transposon region encompassed between the 32 bp TIRs, including both external TSDs, and added two nucleotides, TC sequence.

This virtual ancestral sequence, built from *T. durum,* was used as template for a BLAST search in NCBI and PHYTOZOME databases. Interesting results were found in *Arabidopsis*. The *DREB2A* gene in *Arabidopsis thaliana* is located in Chromosome 5 (NCBI ID: AB016570 and consists of one intron and two exons and does not follow the GT-AG rule splice site. Furthermore, the final gene product consists of just the second exon translation. Beside the expected high homology between the AP2 domain of *Arabidopsis* and exon 4 of *DRF1* gene, an interesting similarity was found between 5'UTRs, 3'UTRs and Intron1 of *Arabidopsis* and the corresponding 5'UTRs, 3'UTRs and the virtual ancestral sequences (data not shown). The average score is about 55%

and the observed transition to transversion frequency is 2.44, suggesting that silent substitutions are predominant.

The overall results suggest that the virtual sequence of *DRF1* gene retains a relationship with intron 1 of *DREB2,* despite of the gene evolution. Thus, it is reasonable to assume that both them share a common ancestor and evolved separately after divergence between monocotyledons and eudicotyledons, in Magnoliophyta.

IV – Conclusions

The transposon located inside the *DRF1* gene was studied inside *Triticeae* tribe and in *Brachipodium distachyon*. Our results reflect the taxonomic relationships and accordingly cluster the sequences.

However, we were not able to find reliable relics of a possible ancestor of the gene in the current analysed sequences, even if *Arabidopsis* shares interesting features. More work is necessary to better understand the recovery traces of the past through the evolution.

Acknowledgments

Thiyagarajan Karthikeyan is thankful for a grant from ENEA International Foreign Fellowship in 2012-2014. Furthermore, he was supported by SIGA for attending the Durum Wheat International Symposium (DWIS), in Rome 2013.

References

Chenais B., Caruso A., Hiard S., Casse N., 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments*. Gene,* 509(1), pp. 7–15.

Fedoroff N., 2000. Transposons and genome evolution in plants. *Proc. Nat. Acad. Sci. USA,* 97(13), pp. 7002-7007**.**

Karthikeyan T., Latini A., Galeffi P., Porceddu E., 2009. A non-autonomous DNA transposon inserted in the first and third intron of the *AsDRF1* gene. *Repbase Reports,* Vol. 9(3), pp. 726**.**

Latini A., Rasi C., Sperandei M., Cantale C., Iannetta M., Dettori M., Ammar K., Galeffi P., 2007. Identification of a DREB-related gene in *Triticum durum* and its expression under water stress conditions. *Annals Appl. Biology,* 150(2), pp. 187-195.

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., Bennetzen J.L., 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science,* 274, pp. 765–768.

Sela N., Kim E., Ast G., 2010. The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biology,* 11, R59. doi: 10.1186/gb-2010-11-6-r59. Epub 2010 Jun 2

Sorek R., 2007. The birth of new exons: Mechanisms and evolutionary consequences. *RNA,* 13, pp. 1603–1608.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum L kelihood, Evolutionary Distance and Maximum Parsimony Methods. *Mol. Biol. Evol.,* 28, pp. 2731-2739.

Thompson J.D., Higgins D.G., Gibson T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., 22, pp. 4673–4680.