ACTA ABP BIOCHIMICA POLONICA

Vol. 65, No 3/2018
391–396
https://doi.org/10.18388/abp.2017_1590

Regular paper

Transposon-associated polymorphisms of stress-responsive gene promoters in selected accessions of *Arabidopsis thaliana*

Mladen Naydenov[1]✉, Nadezhda Gospodinova[1], Elena Apostolova[1], Nikolay Anachkov[1], Vesselin Baev[1,2], Mariyana Gozmanova[1,2], Ivan Minkov[2] and Galina Yahubyan[1,2]

[1]Department of Plant Physiology and Molecular Biology, University of Plovdiv, Plovdiv, Bulgaria; [2]Institute of Molecular Biology and Biotechnology, Plovdiv, Bulgaria

Genetic diversity caused by transposable element movement can play an important role in plant adaptation to local environments. Regarding genes, transposon-induced alleles were mostly related to gene bodies and a few of them to promoter regions. In this study, promoter regions of 9 stress-related genes were searched for transposable element insertions in 12 natural accessions of *Arabidopsis thaliana*. The promoter screening was performed *via* PCR amplification with primers designed to flank transposable element insertions in the promoter regions of the reference accession Col-0. Transposable element-associated insertion/deletion (indel) polymorphisms were identified in 7 of the 12 promoter loci across studied accessions that can be developed further as molecular markers. The transposable element absence in the promoter regions of orthologous genes in *A. lyrata* indicated that the insertion of these transposable elements in *A. thaliana* lineage had occurred after its divergence from *A. lyrata*. Sequence analysis of the promoter regions of *CML41* (Calmodulin-like protein 41) and *CHAP* (chaperone protein dnaJ-related) confirmed the indel polymorphic sites in four accessions – Col-0, Wassilewskija, Shahdara, and Pirin. The observed indel polymorphism of the *CHAP* promoter region was associated with specific gene expression profiles in the different accessions grown at a normal and elevated temperature in a plant growth chamber. The collected data can be a starting point for gene expression profiling studies under conditions resembling the natural habitats of accessions.

Key words: transposable elements, promoters, natural accessions, *Arabidopsis*

Received: 19 April, 2017; revised: 03 November, 2017; accepted: 21 May, 2018; available on-line: 27 August, 2018

✉e-mail: naydenov@plantgene.eu

Abbreviations: CHAP, chaperone protein dnaJ-related; CML41, Calmodulin-like protein 41; RdDM, RNA-directed DNA methylation; TE, Transposable element

INTRODUCTION

Transposable elements (TEs) make up a substantial proportion of plant genomes (Lisch, 2009). Three major classes of TE have been characterized by now. The class-I elements, or retrotransposons, transpose *via* an RNA intermediate by a "copy-and-paste" mechanism (Feschotte *et al*., 2002). This class includes SINEs (short interspersed nuclear elements), LINEs (long interspersed nuclear elements) and LTR retrotransposons which are flanked by long terminal repeats (LTRs) (Kazazian, 2004; Schmidt, 1999). The class-II elements, or DNA TEs, employ a "cut and paste" strategy for transposition without the use of an intermediate (Feschotte *et al*., 2002). Helitrons, the third distinctive class of TE, are hypothesized to transpose *via* a "rolling circle" mechanism (Kapitonov & Jurka, 2001; Kidwell, 2002).

TE transposition can interfere with the gene function (Feschotte & Pritham, 2007; Lisch, 2009), and pose a significant threat to genome stability and integrity (Goettel & Messing, 2009). Therefore, the TEs which have resided in a genome over a long period of evolutionary time are likely to co-evolve with their host genomes to prevent serious disruption of the host gene activity and fitness. Plants have evolved a unique mechanism to attenuate TE activity known as RNA-directed DNA methylation, or RdDM (Mosher *et al*., 2008; Zilberman *et al*., 2007). Extensive molecular studies of *Arabidopsis* and rice have revealed that TE inactivation through the RdDM pathway is associated with DNA methylation, targeted by 24-nt small interfering RNAs, or siRNAs (Kasschau *et al*., 2007; Lister *et al*., 2008; Pontier *et al*., 2005; Zilberman *et al*., 2007).

The altered control of TE activation may affect genes that are important for plant development and stress response (Chinnusamy & Zhu, 2009; Ito *et al*., 2013). It was shown that the RdDM-mediated silencing of SUPPRESSOR OF drm1 drm2 cmt3 (SDC) promoter is a necessary process for normal leaf development (Henderson & Jacobsen, 2008). Stress can change gene expression through DNA methylation and may induce TE activation and movement (Grandbastien, 2004; Ito *et al*., 2013). For example, cold stress-induced hypomethylation triggers transposition of the Tam-3 transposon in *Antirrhinum majus* (Hashida *et al*., 2006), siRNAs and DNA methylation were shown to be associated with the Tnt1 transposon in *Solanaceae* (Andika *et al*., 2006). Differential expression of both, endogenes and transgenes in response to stress can be regulated by RdDM of TEs which reside in gene promoter regions (Grandbastien, 2004; Kashkush *et al*., 2003; Steward *et al*., 2002). Loss of TE silencing in response to stress may increase phenotypic diversity partially through novel TE insertion (Ito, 2012), which in turn might increase adaptability of plants to changing environment (Matzke *et al*., 2015).

Therefore, it is no surprise that genomic analyses of *Arabidopsis thaliana* accessions revealed structural variations in approximately 80% of TEs (Cao *et al*., 2011; Gan *et al*., 2011; Vaughn *et al*., 2007). It was shown that siRNAs are enriched in TE regions, which are present in some accessions but missing in others, and that the siRNA targeting of TEs may promote sequence deletions from the genome bringing about diversification of gene

expression in plants (Wang *et al*., 2013). In our previous work, we determined a number of promoter regions of stress-responsive genes which contain TE insertions and are potential targets for 24 nt siRNAs and DNA methylation in the accession Columbia-0 (Col-0) of *A. thaliana* (Baev *et al*., 2010). Here, we searched the promoter regions of 9 genes related to abiotic and biotic stress for TE insertions, in 12 natural accessions of *A. thaliana*. Of the genes showing TE-associated polymorphism in their promoters, two genes: *CML41* and *CHAP* were further analyzed. The polymorphic promoter regions were sequenced and the associated *cis*-regulatory elements were uncovered in Col-0, Wassilewskija, Shahdara and Pirin accessions. The *CHAP* transcript levels were assessed in the four accessions grown under laboratory conditions at 21°C and 36°C.

MATERIALS AND METHODS

Plant material and high-temperature treatment. Seeds of *Arabidopsis thaliana* natural accessions were bought from the Nottingham Arabidopsis Stock Centre (NASC): Col-0 (with NASC ID N1092), Wassilewskija (Ws-0, N1602), Landsberg (Ler-1, N1642), Shahdara (Shah, N6180), C24 (N22620), N2 (N22480), Nok-0 (N1398), Pf-0 (N1452), Cal-0 (N1062), Cvi-0 (N1096), Yo-0 (N1622), and Pirin (Pirin-12, N28882). They were planted in soil and stratified at 4°C for 5 days. Seedlings were grown in a Percival chamber at 21°C, 16 h light (100 umol m^{-2} s^{-1}) /8 h dark for 36 days, at relative humidity of 50–65%. For high-temperature treatment, plants were transferred to 36°C, 16 h light (100 umol m^{-2} s^{-1}) /8 h dark light cycle, at relative humidity of 15–25%. Rosette leaves of four plants from each accession were taken at 0, 6, 24 and 48 h following the high-temperature treatment.

Total RNA and DNA extraction. Total RNA was extracted using RNeasy Plant Mini Kit (Qiagen) and treated with DNase (Qiagen) according to the manufacturer's instructions. RNA samples from all time points were used for cDNA synthesis and qRT-PCR expression analysis. Total DNA was extracted using DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions.

PCR. For PCR amplification, PCR mix consisted of 0.5 μl DNA, 1 μl dNTPs (5 mM), 2.5 μl Buffer (10x), 1.5 μl $MgCl_2$ (25 mM), 0.5 μl Fw/Rev Primer (10 μM), 0.13 μl Taq polymerase (5 U/μl) and Nuclease-free H_20 added to a final volume of 25 μl per reaction. PCR amplification was performed under the following conditions: 94°C for 4 min, followed by 30 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 45 s to 2 min (depending on the fragment length), and final elongation at 72°C for 7 min.

qRT-PCR. For cDNA synthesis: 1 μg of RNA was reverse transcribed with RevertAid Reverse Transcriptase Kit (Thermo Scientific) following the manufacturer's instructions. For subsequent expression analysis cDNA was diluted 20 times with nucleases-free water. For amplification: PCR amplification was performed using a standard SYBR Green protocol (Fermentas) in a 7500 Real-time PCR machine (Applied Biosystems). All reactions were carried out in a total volume of 25 μl and contained 5 μl of diluted cDNA, 1.5 μl of primer mix at a final concentration of 0.6 mM, 4.5 μl of Nuclease-free water and 12.5 μl of SYBR Green mix with ROX. PCR conditions: 50°C for 2 min, 95°C for 10 min, 40 amplification cycles of 95°C for 15 s and 60°C for 1 min. All reactions were performed in triplicates. The housekeeping gene – EF1a was used as an endogenous control for normalization (Czechowski *et al*., 2005), and the untreated sample was accepted as a reference (RQ j 1). Ct values were calculated using the 7500 software v.2.0.1 (ABI). Relative quantitation of gene expression (RQ) was determined with the equation: E to the power of –ΔCt of gene of interest/E to the power of –ΔCt of housekeeping gene, where E is the primer efficiency measured by standard curve experiment with serial dilutions, and ΔCt is the difference between the Ct values of target gene for each sample and for the reference sample. The primer sequences were designed to match gene conservative regions which had been determined using the Polymorph Variant browser of the 1001 Genomes Project with uploading all available accessions (Supplemental file 1 at www.actabp.pl).

Cloning and sequencing. For ligation: PCR products were cloned into the pTZ57R/T vector (Thermo Scientific) according to the manufacturer's instructions, and sequenced by LGC Genomics, Germany.

Software and web based analysis tools. The promoter regions of *A. thaliana* (Col-0) of analyzed genes were retrieved from the Arabidopsis Gene Regulatory Information Server (AGRIS, www.arabidopsis.med.ohio-state.edu/AtcisDB/) (Davuluri *et al*., 2003). In order to identify the orthologous genes in *A. lyrata,* the amino acid sequences of *A. thaliana* genes extracted from Arabidopsis Information Resource (TAIR) (www.arabidopsis.org) were used in the TBLASTN (protein query to translated 6 frames nt db) tool of Phytozome (V11.0, www.phytozome.net) in the *A. lyrata* genome. 2000 bp upstream of the start codon were considered as putative promoter regions in *A. lyrata*. Sequences were formatted as multi-FASTA formats and were intersected via accession number with the TAIR10 (or PHYTOZOME) gene functional description file. A local copy of RepeatMasker (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015, www.repeatmasker.org) was used to identify and classify TE fragments in promoter sequences. Sequence data of *cis*-regulatory elements was retrieved from AGRIS. The phylogenetic tree was constructed on a data matrix based on the presence (1) or absence (0) of a polymorphic region in the promoter of analyzed genes. Data were statistically analyzed by the FreeTree software program which computed the distance matrix and constructed the phylogenetic tree using the unweight pair group arithmetic average-linkage algorithm. The tree was visualized by FigTree v1.4.2.

RESULTS

TE-determined indel polymorphism of promoter regions in different *Arabidopsis* accessions

In this study, the promoter regions of 9 stress-responsive genes – At1g02450 (NIM1-INTERACTING 1, NIMIN1), At1g14790 (RNA-dependent RNA polymerase 1, RDR1), At3g12500 (Pathogen-related 3, PR3), At3g29810 (COBRA-like protein 2, COBL2), At3g50770 (Calmodulin-like protein 41, CML41), At4g08390 (chloroplastic stromal ascorbate peroxidase, sAPX), At4g11600 (Glutathione peroxidase 6, GPX6), At4g09460 (ATMYB6) and At5g43260 (chaperone protein dnaJ-related, CHAP) were tested for the presence of TE in 12 accessions of *A. thaliana,* of which 10 originated from specific environments. The examined promoters were chosen on the basis of the analysis previously done by Baev and coworkers (Baev *et al*., 2010). Accord-

Table 1. TE-associated indel polymorphisms of the promoter regions of nine stress-responsive genes in different accessions of *A. thaliana*.
TE fragments in the reference genome Col-0 and in *A. lyrata* were detected and classified by RepeatMasker (version 4.0.6). In accordance with the PCR amplification data presented on Fig. 1, N stands for a non-polymorphic site and P stands for a polymorphic site when compared to the respective Col-0 site. The transposable element length is given in brackets.

Locus ID	Col-0 (TE)	Ws	Ler-1	C24	Shah	N2	Noc-0	Pf-0	Cal-0	Cvi-0	Yo-0	Pirin	lyrata
At1g02450	RC/Helitron (1997nt)	N	N	N	N	N	N	N	N	N	N	N	DNA/Tc-Mar-Pogo (454 nt)
At1g14790	DNA/MULE-MuDR (299 nt)	N	N	N	N	N	N	N	N	N	N	N	N
At3g12500	RC/Helitron (1049 nt)	N	N	N	P	P	N	N	P	N	N	N	N
At3g29810	DNA/hAT (1194 nt) DNA/MULE-MuDR (285nt)	P	P	P	P	P	P	N	P	P	P	P	N
At3g50770	DNA (279 nt)	P	P	P	P	P	P	N	P	P	P	P	N
At4g08390	DNA/MULE-MuDR (279 nt)	N	N	N	N	N	N	N	N	N	N	N	N
At4g11600	DNA/MULE-MuDR (355 nt and 112 nt)	P	N	N	P	N	N	N	P	N	N	N	N
At4g09460	LTR/Copia (2013 nt), DNA/MULE-MuDR (625 nt)	P	P	P	P	P	P	P	P	P	P	P	LTR/Copia (447 nt)
At5g43260	RC/Helitron (89 nt and 32 nt)	P	P	N	P	P	P	P	P	P	P	P	N

ing to it, each of these promoters contains a TE which can be a site for methylation and siRNA targeting in the Col-0 accession's genome. Here, the putative promoter regions of orthologous genes in *A. lyrata* were searched for TE-derived sequences by RepeatMasker. TEs were only found in the promoters of the genes orthologous to At1g02450 and At4g09460, but they were different from those identified in the respective Col-0 promoters (Table 1).

To assess the TE-based differences of promoter regions in the analyzed accessions of *A. thaliana*, primer pairs were designed to flank the TE boundaries in the reference accession Col-0. Because only the Col-0 genome has been completely sequenced and annotated, it was chosen as a reference genome and the deletion or insertion mutations were defined according to it.

PCR amplification showed fragments of different size for a particular promoter region in different accessions (Fig. 1). Since primers were designed to flank the TEs in the reference promoter sequences of Col-0, the observed polymorphic products should reflect a TE absence from the corresponding promoter region of a particular accession when compared to Col-0. Insertion/deletion (indel) polymorphism was identified in the promoter regions of At1g02450, At3g12500, At3g29810, At3g50770, At4g11600 and At5g43260, and was specific for each accession (Fig. 1, Table 1). Notably, the length of missing fragments was approximately the same as that of the corresponding TE in Col-0. For At4g09460, no amplification of the promoter region was observed in all accessions except for Col-0. Since this result was observed with two different primer pairs, the lack of amplified products may imply severe changes of the region in the analyzed accessions. No indel polymorphism was identified in the promoter regions of At1g14790 and At4g08390.

The described polymorphic regions were used for construction of a phylogenetic tree which showed that the 12 *A. thaliana* accessions could be divided into three major groups (Fig. 2). The largest group was comprised of Ws, C24, Pirin, Yo-0, Cvi-0, Ler-1 and Noc 0, and shared a closer relationship with the group of Shah, Cal-0 and N2. Col-0 and Pf-0 were less closely related to the other accessions.

Sequence analysis of the CML41 and CHAP promoters in Col-0, Ws, Shah, and Pirin accessions

Sequence analysis of the indel polymorphic sites in the promoter regions of *CML41* and *CHAP* was performed in the four accessions Col-0, Ws, Shah, and Pirin. The two promoter regions were chosen because we had previously found that they were both under the control of the RdDM pathway (Baev *et al.*, 2010). The comparative analysis of the sequenced fragments, produced with the primer pairs described in the previous section, confirmed the polymorphism between Col-0 and the other three accessions (Supplemental file 2 at www.actabp.pl, Fig. 3). In Col-0, the *CML41* promoter region contains an insertion of 302 bp which is comprised of a DNA TE from the RP1_AT family and the *CHAP* promoter region contains an insertion of 789 bp comprised of two ATREP7 remnants. The sequence alignments revealed high homology between the analyzed regions in Ws, Shah, and Pirin.

To find *cis*-regulatory elements in the polymorphic insertions in the *CML41* and *CHAP* promoters of Col-0, the cognate sequences were searched for known *cis*-elements from the AGRIS plant database (Fig. 3). The insertion in the *CML41* promoter contains 6 *cis*-regulatory motifs – GATA, Bellringer/replumless/pennywise BS1 IN AG, SORLIP2, SORLIP3. Nine *cis*-regulatory elements (AtMYC2 BS in RD22, 2 RAV1-A binding site motif, 2 Bellringer/replumless/pennywise BS1 IN AG, GATA, ARF1 binding site motif, L1-box, ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH) were detected in the insertion in the *CHAP* promoter.

Intraspecific variation of CHAP expression in Col-0, Ws, Shah, and Pirin accessions

In our previous study we observed that the *CHAP* expression was activated in response to elevated tempera-

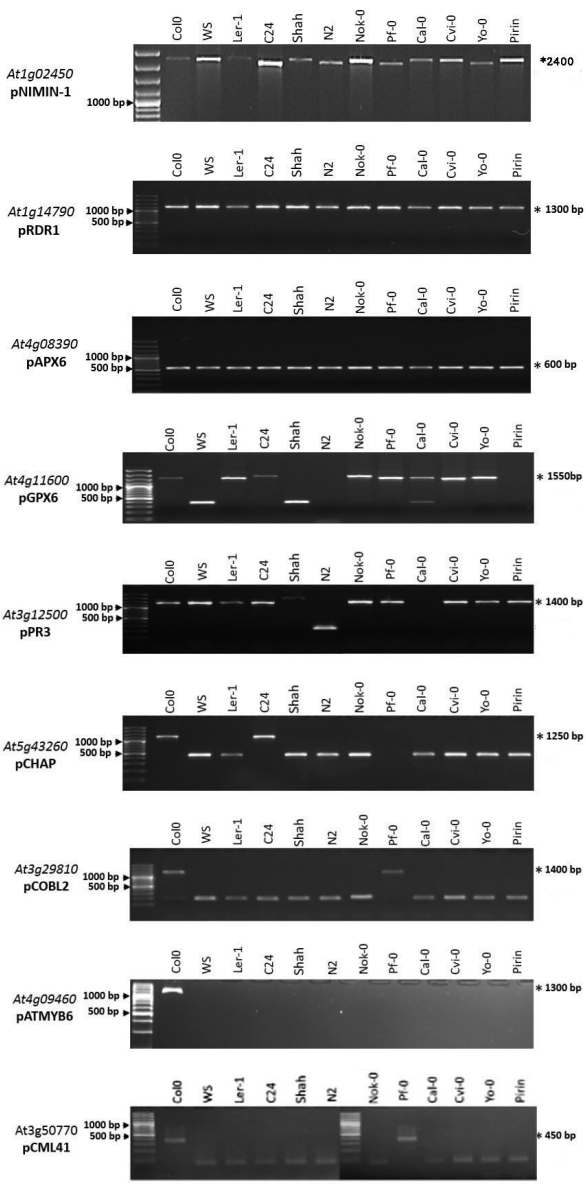

Figure 1. TE-associated indel polymorphisms of the promoter regions of nine stress-responsive genes in different accessions of *A. thaliana*.
PCR amplification was performed with primers designed to flank TE insertions in the respective promoter regions of the reference Col-0 accession. The size of the expected PCR fragments is indicated with an asterisk in Col-0.

ture (Naydenov *et al.*, 2015). In order to answer whether the observed indel polymorphism of the *CHAP* promoter region in different accessions could affect its stress response, the four ecotypes were treated with high temperature for 48 h. The transcript levels of *CHAP* were lower in Ws and Pirin and higher in Shah compared to Col-0 when the plants were grown at 21°C (Fig. 4). The high-temperature treatment resulted in an increased expression of *CHAP* in Col-0 and affected gene expression in a specific manner in the other three accessions. The elevated temperature caused a gradual increase of the *CHAP* expression in Ws and Pirin, while the transcript level decreased in the first hours of treatment in Shah and reached the highest values at 48 h of treatment in comparison with the other accessions (Fig. 4).

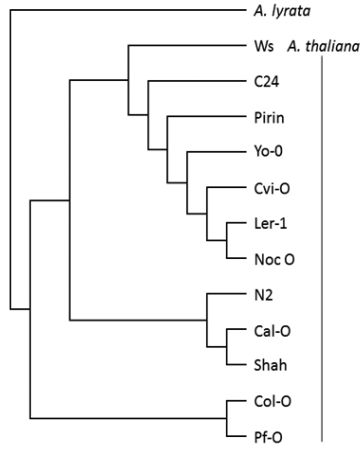

Figure 2. Phylogenetic tree of 12 *A. thaliana* accessions based on the TE-associated polymorphic promoter regions.
The tree was constructed by application of Nei and Li/Dice similarity index and UPGMA clustering method.

DISCUSSION

A great number of small-scale polymorphisms, as well as many larger insertions and deletions, have been described in the genomes of numerous *A. thaliana* accessions by means of SNP chips and NGS, and further subjected to genome-wide association (GWA) studies (Cao *et al.*, 2011; Schneeberger *et al.*, 2011). TE indel polymorphism has been associated with plant adaptation to local environments (Casacuberta & Gonzalez, 2013), for example to diverse light conditions in *Arabidopsis* (Lin *et al.*, 2007). By analyzing genes, the TE-based polymorphisms were located mostly within gene bodies, and few of them within promoter regions (Muterko *et al.*, 2015).

Genome- and transcriptome-wide studies have revealed that TE fixation may be prevented within a population through negative selection, especially for those that are close to genes (Hollister & Gaut, 2009; Wang *et al.*, 2013). Our study reveals TE presence/absence polymorphisms in the promoter regions of seven stress-responsive genes in *A. thaliana* accessions, originating from various geographical locations. The products of these genes (CML 41, CHAP, GPX6, MYB6, PR3 protein) have various functions in plant cell defense. Notably, the length of a particular TE does not differ across the accessions in which this TE is present. Unlike the genes discussed above, the promoter regions of the genes encoding RDR1 and sAPX6 have retained TEs of the MULE-MuDR family across all accessions.

The observed indel polymorphism demonstrates that TE activity can be affected by environmental pressure and can alter the set of regulatory elements next to genes by bringing new promoter motifs or purging some of the existing ones. The sequencing of the promoter regions of *CHAP* and *CML41* from Col-0, Ws, Shah, and Pirin reveals that the TE-associated insertions in Col-0 have been a source of a number of new regulatory elements for the regulatory areas of the two genes. For example, two RAV1 promoter motifs were found in the polymorphic region of the *CHAP* promoter of Col-0 that could bind the RAV1 transcription factor, which may have roles in growth and stress responses (Hu *et al.*, 2004; Yamasaki, *et al.*, 2004).

The TE absence in the promoter regions of orthologous genes in *A. lyrata* indicates that the insertion of the analyzed TEs in *A. thaliana* lineage has occurred after its

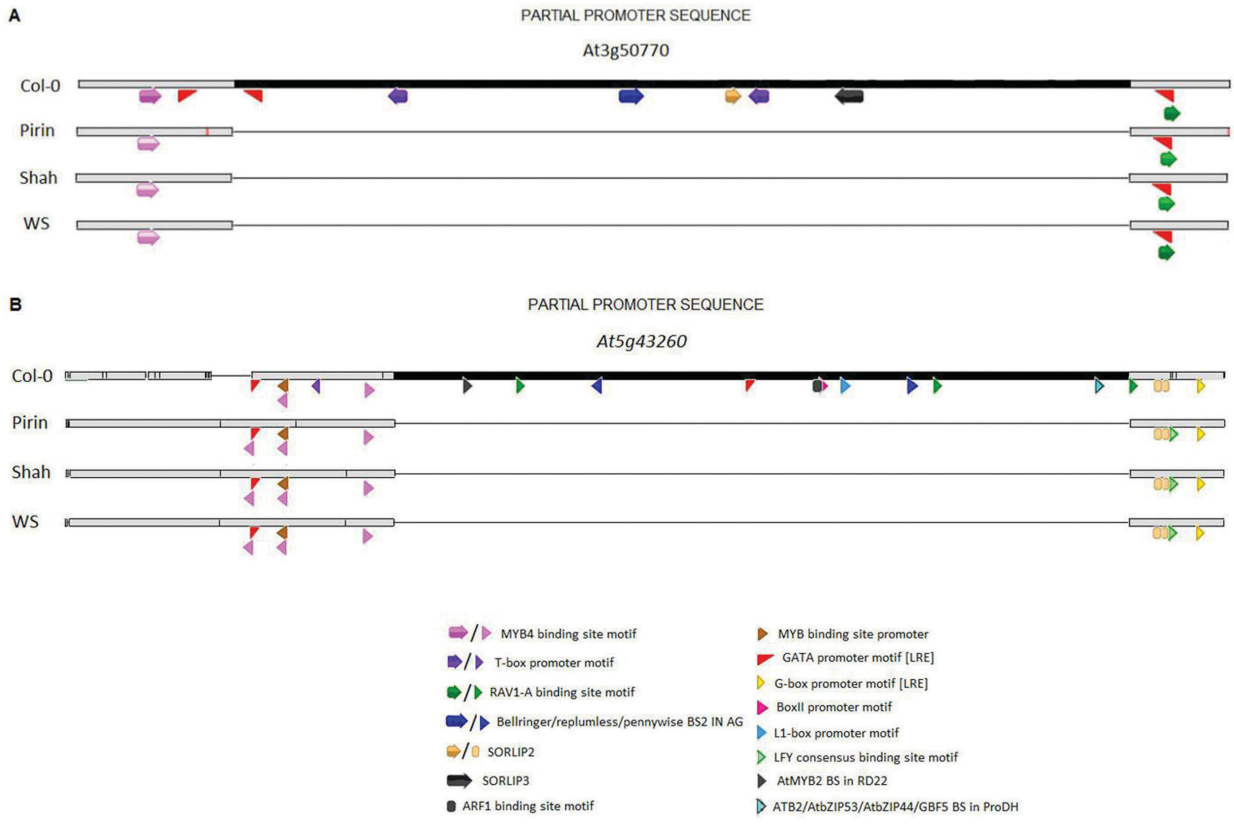

Figure 3. Polymorphic regions in the *CHAP* and *CML41* promoters in the Col-0, Ws, Shah, and Pirin accessions identified with direct sequencing.
The TE-associated insertions in the two promoters in Col-0 are represented with a solid black line. Gaps inserted during the sequence alignments in the promoter regions of Ws, Shah, and Pirin are represented with a thin black line. *Cis*-regulatory elements were from the Arabidopsis Gene Regulatory Information Server (AGRIS).

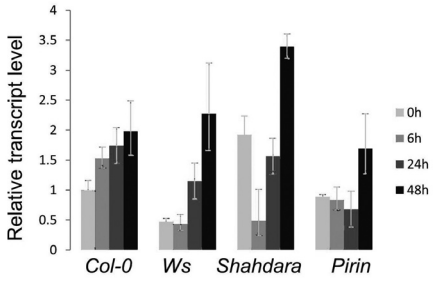

Figure 4. Expression profiles of *CHAP* in the Col-0, Ws, Shan, and Pirin accessions.
The Relative Quantity (RQ) was determined using Col-0 plants (0h) grown at 21°C as a reference sample. The average results of three independent experiments, with the standard deviation of the mean, are represented on the histograms.

divergence from *A. lyrata*. The Col-0 is clustered with the Pf-0 accession (originated from West Germany) in a separate subgroup, providing evidence that the true origins of Col-0 could be traced to West Germany (http://www.lehleseeds.com). The reported indel polymorphisms in *Arabidopsis* gene promoters can be selected for genetic markers in the same way as the MITE-related genetic markers in the 3' regions of maize genes (Bhattramakki *et al.*, 2002).

One of the ways in which TE may affect the function of neighboring genes is *via* altering promoter sequences that in turn can have negative or positive consequences on gene expression (Kashkush *et al.*, 2003; Zhang *et al.*, 2008). *CHAP* illustrates this well – we observed that the TE polymorphism of the *CHAP* promoter resulted in gene expression divergence between orthologous genes in the four analyzed accessions – Col-0, Ws, Shah, and Pirin. Col-0 responded to temperature stress with a gradual increase in *CHAP* expression whereas in the other three accessions the gene activation was delayed and found 48 hours after application of the stress factor. The insertion of the ATREP7 remnants in the *CHAP* promoter region in Col-0 is an example of the transposon ability to cause huge structural changes in promoter regions with strong regulatory effects stemming from the acquisition of new regulatory elements carried by the transposon itself.

In conclusion, our results confirm that TEs may take part in the diversification of the regulatory regions such as the promoters in plant genomes in order to adjust the gene expression of natural accessions to specific environments. Gene expression control is a complex process in which *cis*-acting promoter elements interplay with *trans*-acting factors (proteins and small RNAs) to mediate DNA and chromatin modifications, and is influenced by a wide variety of environmental factors. Because of these reasons, additional experiments, especially under conditions resembling natural habitats, are needed to elucidate the impact of a particular TE-induced promoter polymorphism on the species adaptation and evolution. Moreover, the identified TE-associated indel polymorphisms can be developed further as molecular markers for distinguishing the natural populations of *A. thaliana*.

Acknowledgements

This work was partially supported by the Bulgarian Ministry of Education and Sciences, Grant DH06/6 and SA 739582 – PlantaSYST, H2020 – Widespread-2016-2017-Teaming Phase 2.

REFERENCES

Andika IB, Kondo H, Rahim MD, Tamada T (2006) Lower levels of transgene silencing in roots is associated with reduced DNA methylation levels at non-symmetrical sites but not at symmetrical sites. *Plant Mol Biol* **60**: 423–35. doi: 10.1007/s11103-005-4429-7

Baev V, Naydenov M, Apostolova E, Ivanova D, Doncheva S, Minkov I, Yahubyan G (2010) Identification of RNA-dependent DNA-methylation regulated promoters in *Arabidopsis*. *Plant Physiol Biochem* **48**: 393-400. doi: 10.1016/j.plaphy.2010.03.013

Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC, Tingey SV, Rafalski A (2002) Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Mol Biol* **48**: 539–547. doi: 10.1023/A:1014841612043

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D. (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* **43**: 956–963. doi: 10.1038/ng.911

Casacuberta E, Gonzalez J (2013) The impact of transposable elements in environmental adaptation. *Mol Ecol* **22**: 1503–1517. doi: 10.1111/mec.12170

ChinnusamyV, Zhu J K (2009) Epigenetic regulation of stress responses in plants. *Curr Opin Plant Biol* **12**: 133-139. doi: 10.1016/j.pbi.2008.12.006

Chu CG, Tan CT, Yu G T, Zhong S, Xu SS,Yan L (2011) A Novel Retrotransposon inserted in the dominant vrn-b1 allele confers spring growth habit in tetraploid wheat (*Triticum turgidum* L.). *G3 (Bethesda)* **1**: 637–645. doi: 10.1534/g3.111.001131

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol* **139**: 5–17. doi: 10.1104/pp.105.063743

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors. *BMC Bioinformatics* **4**: 25. doi: 10.1186/1471-2105-4-25

Feschotte C, Jiang N,Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3: 329-41. doi: 10.1038/nrg793

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**: 331–368. doi: 10.1146/annurev.genet.40.110405.090448

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, Kahles A, Bohnert R, Jean G, Derwent P, Kersey P, Belfield EJ, Harberd NP, Kemen E, Toomajian C, Kover PX, Clark RM, Rätsch G, Mott R. (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature* **477**: 419–423. doi: 10.1038/nature10414

Goettel W, Messing J (2009) Change of gene structure and function by non-homologous end-joining, homologous recombination, and transposition of DNA. *PLoS Genet* **5**: e1000516. doi: 10.1371/journal.pgen.1000516

Grandbastien M A (2004) Stress activation and genomic impact of plant retrotransposons. *J Soc Biol* **198**: 425–432

Hashida SN, Uchiyama T, Martin C, Kishima Y, Sano Y, Mikami T (2006) The temperature-dependent change in methylation of the Antirrhinum transposon Tam3 is controlled by the activity of its transposase. *Plant Cell* **18**: 104–118. doi: 10.1105/tpc.105.037655

Henderson IR., Jacobsen SE (2008) Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* **22**: 1597–1606. doi: 10.1101/gad.1667808

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**: 1419–1428. doi: 10.1101/gr.091678.109

Hu YX, Wang YH, Liu XF, Li JY (2004) *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development *Cell Res* **14**: 8–15. doi: 10.1038/sj.cr.7290197

Ito H (2012) Small RNAs and transposon silencing in plants. *Dev Growth Differ* **54**: 100–107. doi: 10.1111/j.1440-169X.2011.01309.x

Ito H, Yoshida T, Tsukahara S, Kawabe A (2013) Evolution of the ONSEN retrotransposon family activated upon heat stress in Brassicaceae. *Gene* **518**: 256–261. doi: 10.1016/j.gene.2013.01.034

Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A* **98**: 8714–8719. doi: 10.1073/pnas.151269298

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet* **33**: 102–106. doi: 10.1038/ng1063

Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC (2007) Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol* **5**: e57. doi: 10.1371/journal.pbio.0050057

Kazazian HH, Jr (2004) Mobile elements: drivers of genome evolution. *Science* **303**: 1626–1632. doi: 10.1126/science.1089670

Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**: 49–63. doi: 10.1023/A:1016072014259

Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H (2007) Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. *Science* **318**: 1302–1305. doi: 10.1126/science.1146281

Lisch D (2009) Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* **60**: 43–66. doi: 10.1146/annurev.arplant.59.032607.092744

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536. doi: 10.1016/j.cell.2008.03.029

Matzke MA, Kanno T, Matzke AJ (2015) RNA-Directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annu Rev Plant Biol* **66**: 243–267. doi: 10.1146/annurev-arplant-043014-114633

Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, Mayer KF, Wing RA (2004) Sequence composition and genome organization of maize. *Proc Natl Acad Sci U S A* **101**: 14349–14354. doi: 10.1073/pnas.0406163101

Mosher RA, Schwach F, Studholme D, Baulcombe DC (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci U S A* **105**: 3145–3150. doi: 10.1073/pnas.0709632105

Muterko A, Kalendar R, Cockram J, Balashova I (2015) Discovery, evaluation and distribution of haplotypes and new alleles of the Photoperiod-A1 gene in wheat. *Plant Mol Biol* **88**: 149–164. doi: 10.1007/s11103-015-0313-2

Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev* **19**: 2030–2040. doi: 10.1101/gad.348405

Schmidt T (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Mol Biol* **40**: 903–910. doi: 10.1023/A:1006212929794

Schneeberger K, Ossowski S, Ott F, Klein JD, Wang X, Lanz C, Smith LM, Cao J, Fitz J, Warthmann N, Henz SR, Huson DH, Weigel D (2011) Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc Natl Acad Sci U S A* **108**: 10249–10254. doi: 10.1073/pnas.1107739108

Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H (2002) Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *J Biol Chem* **277**: 37741–37746. doi: 10.1074/jbc.M204050200

Yamasaki K1, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S. (2004) Solution structure of the B3 DNA binding domain of the *Arabidopsis* cold-responsive transcription factor RAV1. *Plant Cell* **16**: 3448–3459. doi: 10.1105/tpc.104.026112

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW, Martienssen RA. (2007) Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol* **5**: e174. doi: 10.1371/journal.pbio.0050174

Wang X, Weigel D, Smith LM (2013) Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genet* **9**: e1003255. doi: 10.1371/journal.pgen.1003255

Zhang HY, He H, Chen LB, Li L, Liang MZ, Wang XF, Liu XG, He GM, Chen RS, Ma LG, Deng XW (2008) A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Mol Plant* **1**: 720–731. doi: 10.1093/mp/ssn022

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61–69. doi: 10.1038/ng1929