Caryologia
International Journal of Cytology, Cytosystematics and Cytogenetics

Firenze University Press
www.fupress.com/caryologia

OPEN ACCESS

Citation: A. Vangelisti, G. Usai, F. Mascagni, L. Natali, T. Giordani, A. Cavallini (2019) A whole genome analysis of long-terminal-repeat retrotransposon transcription in leaves of *Populus trichocarpa* L. subjected to different stresses. *Caryologia* 72(2): 69-79. doi: 10.13128/cayologia-232

Published: December 5, 2019

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

A whole genome analysis of long-terminal-repeat retrotransposon transcription in leaves of *Populus trichocarpa* L. subjected to different stresses

Alberto Vangelisti[#], Gabriele Usai[#], Flavia Mascagni[#], Lucia Natali, Tommaso Giordani[*], Andrea Cavallini

Department of Agriculture, Food and Environment, University of Pisa, Via del Borghetto 80, I-56124 Pisa, Italy

[#] These authors contributed equally to this study

[*] Corresponding author: tommaso.giordani@unipi.it

Abstract. Long terminal repeat retrotransposons have a main role in shaping the structure of plant genomes. We used available genomic resources to study as several factors affect the expression of long terminal repeat retrotransposons in *Populus trichocarpa*. Such factors included redundancy of a retrotransposon in the genome, chromosomal localization, "genotype" of the retrotransposon, and changes in the environment. Overall, we identified and annotated 828 full-length retrotransposons, and analyzed their abundance in the genome. Then, we measured their expression in leaves of plants subjected to several stresses (drought, cold, heat, and salt) as well as in control plants. Our analyses showed that the expression of retrotransposons was generally low, especially that of abundant elements. The transcription of an element was found to be only slightly dependent on its chromosomal localization, rather it depended on the superfamily and the lineage to which the retrotransposon belonged. Finally, some retrotransposons were specifically activated by different environmental stresses.

Keywords. LTR-retrotransposons, retrotransposon expression, retrotransposon abundance, Illumina cDNA libraries, *Populus trichocarpa*.

INTRODUCTION

Transposable elements are mobile DNA sequences, which are abundant and widespread in all eukaryotic genomes. They can change their position on chromosomes by a mechanism, called transposition, driven by enzymes encoded by the element itself. Transposable elements can be divided between retrotransposons (REs, Class I) and DNA transposons (Class II), according to their transposition mechanism (Wicker et al. 2007).

The transposition of REs occurs through a "copy and paste" replicative mechanism that includes the transcription of an RNA intermediate followed by its retro-transcription and insertion into the genome (Wicker et al.

2007). This transposition mechanism has allowed REs to become the largest portion of most eukaryotic genomes, often represented by many thousands of copies (San-Miguel et al. 1998; Vicient et al. 1999).

A retrotransposon can be classified as LTR- or not LTR-RE, according to the presence of long terminal repeats (LTRs) at its ends. As for the LTR-REs, the promoter elements, the polyadenylation signals and the expression enhancers are found in the LTRs. These domains regulate the transcription of the element (Bennetzen 2000). In the coding portion of the LTR-REs, Gag and Pol domains can be found. Gag encodes virus-like particles, Pol encodes the enzymes necessary to produce new cDNA molecules from the RE transcripts and to integrate them into new sites in the host genome (Bennetzen 2000). Other structural features involved in the RE replication process include a primer binding site and a poly-purine tract (Bennetzen 2000).

The LTR-REs are essentially subdivided into two superfamilies, *Gypsy* and *Copia* (Wicker et al. 2007), according to the order of gene sequences within the Pol domain. Superfamilies, in turn, are distinguished into several lineages in relation to sequence conservation and structure (Barghini et al. 2015a; Usai et al. 2017; Buti et al. 2018; Mascagni et al. 2017; 2018a).

The replicative activity of LTR-REs can determine large variations in the genome structureof eukaryotic species (Springer et al. 2009; Vitte et al. 2014). Among the effects of retrotransposition, besides determining changes in genome size, RE-related structural variations can modify the regulation patterns of protein-encoding genes and, consequently, their activity, influencing the phenotype (Slotkin and Martienssen 2007; Butelli et al. 2012; Falchi et al. 2013; Lisch 2013).

The first phase of retrotransposition is represented by the transcription of the element. The RE transcripts can be capped and polyadenylated or not. In the former case, transcripts should be destined to be translated into the enzymes for retrotransposition, in the latter case, transcripts should be reverse-transcribed (Chang et al. 2013; Meignin et al. 2003).

Transcription of REs has been described in several plant species (Grandbastien 2015). In some grass species LTR-REs are poorly constitutively transcribed (Vicient et al. 2001; Ishiguro et al. 2014). In other species, for example in *Populus* x *canadensis*, certain LTR-REs are expressed constitutively, without apparent induction conditions (Giordani et al. 2016). Retrotransposition is completed when a new copy of the element is inserted into the genome. This has been reported for *Tnt1* and *Tto1* elements of *Nicotiana* and for *Tos17* of rice, induced by tissue culture (Grandbastien 1998). Complete retrotransposition of a *Copia* RE has been described also in sunflower seedlings, grown under standard conditions (Vukich et al. 2009).

Retrotransposition is generally limited by the host genome due to its potentially mutagenic action. A major mechanism to inactivate mobile elements involves the methylation of histones and cytosine residues with consequent silencing of chromatin (Dieguez et al. 1998). Post-transcriptional silencing by RNA degradation also plays an important role in the epigenetic control of RE activity (Slotkin and Martienssen 2007; Lisch 2013; Ito 2013).

In recent years, many studies have been carried on the LTR-REs of the genus *Populus* and in particular on *P. trichocarpa*, which is considered a model species for forest trees. The *P. trichocarpa* genome was the first genome to have been sequenced for a forest species (Tuskan et al. 2006) and has been recently updated (Zeng et al. 2017). This species has a relatively small genome (550 Mbp) and REs cover approximately 176 Mbp (32% of the genome), with a prevalence of *Gypsy* over *Copia* RE sequences (Tuskan et al. 2006). *Populus trichocarpa* REs have been identified and annotated according to their superfamily and lineage, and LTR-RE genomic abundance and age of insertion were analyzed as well (Natali et al. 2015; Mascagni et al. 2018b). *P. trichocarpa* LTR-REs have been also used as a reference for several analyses related to the repetitive component in other species of the genus *Populus* (Giordani et al. 2016; Usai et al. 2017).

The transcription of REs is only the first step for retrotransposition and insertion of new copies of the element in the genome. For this reason, analyses on LTR-RE activity should include searching for new insertion events. However, an overall study of factors potentially able to influence the transcription of these elements is not yet available for poplar. We therefore decided to perform a meta-analysis of LTR-RE expression by using publicly available genomic DNA and cDNA libraries obtained from leaves of plants cultivated under standard conditions or subjected to four types of abiotic stress (cold, drought, heat, and salt). The objectives of this work were to evaluate i) the expression level of REs under standard and stress conditions; ii) the correlation between abundance of REs and their expression level; iii) the possibility that different LTR-REs are induced by different (and specific) stresses; iv) the possibility that the expression of a RE is related to the "genotype" of the RE itself, i.e., to the lineage to which it belongs; v) the possibility that the chromosomal localization of a RE can influence its expression.

METHODS

Isolation of full-length LTR-REs of P. trichocarpa

Putative full-length LTR-REs were isolated from the GCA_000002775.3 version (Zeng et al. 2017) of the *P. trichocarpa* genome sequence (Tuskan et al. 2006; Slavov et al. 2012), deposited at the NCBI site (WGS project number AARH02, http://www.ncbi.nlm.nih.gov/assembly/GCF_000002775.3). Full-length LTR-REs were isolated by using: i) LTRharvest (Ellinghaus et al. 2008) with the following parameters: minlenltr=100, maxlenltr=6000, mindistltr=1500, maxdistltr=25000, mintsd=5, maxtsd=5, similar=85, vic=10, including the presence of TG and CA dinucleotides at 5' and 3'-ends, respectively; ii) LTR-FINDER (Xu et al. 2007), under default.

A random sample of putative LTR-REs (around 20% of the isolated elements) were manually validated using DOTTER (Sonnhammer and Durbin 1995) to verify the occurrence of the two LTRs, of dinucleotides TG and CA at the respective 5' and 3' ends, and of the tandem site duplications. All LTR-REs were annotated by using BLASTN search against plant RE datasets (Barghini et al. 2015b; Natali et al. 2015; Usai et al. 2017; Buti et al. 2018) and by using the Domain Search tool of RepeatExplorer (Novak et al. 2013). Whenever possible, the full-length LTR-REs were identified as belonging to *Gypsy* or *Copia* superfamilies and to the respective lineages.

A multi-FASTA file with the sequences of identified full-length LTR-REs is available at the sequence repository site of the Department of Agriculture, Food and Environment of the University of Pisa (http://pgagl.agr.unipi.it/sequence-repository/).

Illumina cDNA libraries collection

The expression of LTR-REs was analyzed using Illumina cDNA paired-end libraries publicly available at the NCBI SRA (https://www.ncbi.nlm.nih.gov/sra/, BioProject accession PRJEB19784) (Filichkin et al. 2018). Such cDNA libraries were obtained from RNAs from leaves of *P. trichocarpa* (clone Nisqually 1) plants exposed to different stresses, i.e., heat, cold, drought, and salt. All cultivation conditions are described by Filichkin et al. (2018). In brief, for heat stress, plants were treated at 39°C for 12 h (short treatment) or 7 days (prolonged treatment). For cold stress, plants were exposed to cycles of 4°C (night)/12°C (day) for 24 h (short treatment) or 7 days (prolonged treatment). For drought treatment, watering was withheld until soil moisture reached 0.1 m^3/m^3 and maintained at the level of 0.06 –0.1 m^3/m^3 for 5 days (short treatment) or for 12 days after water withholding (prolonged treatment). For salt stress, plants were irrigated with 100 mM NaCl solution for 24 h (short treatment) or for 7 days (prolonged treatment). Three replicate libraries were downloaded for each stress and control plants.

Illumina genomic DNA sequences of the same clone of *P. trichocarpa* were retrieved from the NCBI Sequence Read Archive (NCBI, Washington, USA, https://www.ncbi.nlm.nih.gov/sra, SRA ID SRR1801106).

The quality of the cDNA and genomic DNA reads was checked using FastQC (v. 0.11.3) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the overall quality was improved by removing Illumina adapters and trimming the reads using Trimmomatic (v. 0.38) (Bolger et al., 2014) with different parameters for cDNA (ILLUMINACLIP:2:30:10, SLIDINGWINDOW:4:20, CROP:96, HEADCROP:12 and MINLEN:90) and genomic DNA (ILLUMINACLIP:2:30:10, SLIDINGWINDOW:4:15, CROP:85, MINLEN:85). Organellar sequences were removed from the Illumina libraries by mapping against a database of chloroplast genomes of poplar species (Usai et al. 2017) using CLC-BIO GenomicWorkbench (v. 9.5.3, CLC-BIO, Aarhus, Denmark) with the following parameters: mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.5, similarity fraction 0.8. All matching reads were considered putatively belonging to organellar genomes and removed.

Estimation of retrotransposon expression and abundance in the genome

The expression of LTR-REs was measured mapping cDNA sequence reads of control and cold-, drought-, heat-, or salt-exposed leaves onto the library of *P. trichocarpa* full-length LTR-REs, using CLC-BIO Genomic Workbench with the following parameters: mismatch cost 1, deletion cost 1, insertion cost 1, similarity 0.9 and length fraction 0.9. The expression level of each sequence was calculated and converted both to mapped reads per million (MRpM) and to RPKM (Mortazavi et al. 2008). LTR-REs mapped by 1 to 10 reads per million of reads in at least one sample were considered as expressed (Lu et al. 2013), those mapped by at least 10 reads per million were considered as highly expressed.

Expression values were compared, using Baggerley's test (Baggerley et al. 2003), considering RPKM values in the short and prolonged stage of each treatment in comparison to control leaves. The weighted proportion fold changes between a treatment and controls were considered significant when the weight of a sample was at least two-fold higher or lower than another, with a false dis-

covery rate (FDR; Benjamini and Hochberg, 1995) corrected p-value ≤ 0.05.

In order to assess genomic abundance of REs, genomic DNA reads of *P. trichocarpa* were mapped onto reference retrotransposon domains library using CLC-BIO Genomics Workbench with the same parameters described above. For each LTR-RE the average coverage was calculated. The average coverage is the sum of the bases of the aligned parts of all the reads divided by the length of the reference sequence.

Localization of expressed REs along the poplar genome

Each of the 19 linkage groups (LGs) of the currently available *P. trichocarpa* genome sequence (version GCA_000002775.3, Zeng et al. 2017) were subdivided into 3-Mbp-long genome regions. Then, in order to localize LTR-RE sequences in the genome, the LTR-REs were used for masking the 3-Mbp-long fragments of the poplar genome using RepeatMasker (http://www.repeatmasker.org) with the following parameters: s, no-is, no-low. Masking was performed using i) all isolated full-length elements; ii) all *Chromovirus* LTR-REs; iii) a putative poplar centromeric sequence (Islam-Faridi et al. 2009; Cossu et al. 2012); iv) all LTR-REs expressed in control leaves (mapped by more than ten reads per million). The number of masked bases was then counted for each of the 3 Mbp fragment using an in-house perl script.

RESULTS

Identification of full-length LTR-REs of P. trichocarpa

The full-length LTR-REs used in this study were isolated from the updated genome sequence of *P. trichocarpa* (Zeng et al. 2017), by performing a complete genome scan with LTRharvest and LTRFinder. Besides using these tools with stringent parameters, a sample of isolated elements were manually validated at structural level and all were confirmed as LTR-REs.

The dataset includes 828 full-length LTR-REs. Table 1 reports the number of LTR-REs belonging to the *Gypsy* and *Copia* superfamilies, subdivided according to the lineage to which they belong, i.e. *Athila*, *Ogre* and *Chromovirus* for *Gypsy* elements and *Ale* (distinguished into *AleI* and *AleII*), *Angela*, *Bianca*, *Ivana/Oryco*, *SIRE* and *TAR/Tork* for *Copia* elements. For each lineage the mean average coverage is also reported, calculated after mapping elements with Illumina gDNA reads, which represents the mean abundance of that lineage in the *P. trichocarpa* genome.

Transcription of LTR-REs

The expression of 828 full-length LTR-REs was measured by mapping the elements with Illumina cDNA reads obtained from leaves of plants of *P. trichocarpa* cultivated in standard conditions (controls) and under different stress (drought, heat, cold, or salt). In the control leaves, only 0.47% of the cDNA reads mapped the library, hence, in general, LTR-REs are barely expressed (Fig. 1).

The expression level of LTR-REs decreased with stress, in the decreasing order drought-cold-heat-salt (Fig. 1). No significant difference was observed between short and prolonged treatments, with the exception of cold treatment, where expression decreased reduced in prolonged exposition.

According to Lu et al. (2013), we considered as expressed those LTR-REs mapped by more than one read per million. The number of expressed LTR-REs in controls and in drought-, cold-, heat- and salt-exposed leaves is reported in Fig. 2. The number of LTR-REs expressed in drought-treated leaves is similar to that of control leaves. On the contrary, this number is strongly reduced after the other treatments (Fig. 2). However, the

Table 1. Number and mean average coverage of full-length LTR-REs collected in the *P. trichocarpa* genome (version GCA_000002775.3) and separated according to their superfamily and lineage.

Super-family	Lineage	Nr. of elements	Mean average coverage
Copia	*AleI*	42	14.04
	AleII	122	22.33
	Angela	2	64.13
	Bianca	1	28.24
	Ivana/Oryco	104	19.06
	SIRE	7	42.11
	TAR/Tork	90	17.45
	Total	368	19.89
Gypsy	*Athila*	126	57.37
	Chromovirus	174	40.20
	Ogre	67	41.16
	Unknown	50	13.46
	Total	417	42.34
Unknown		43	22.65
Total		828	31.34

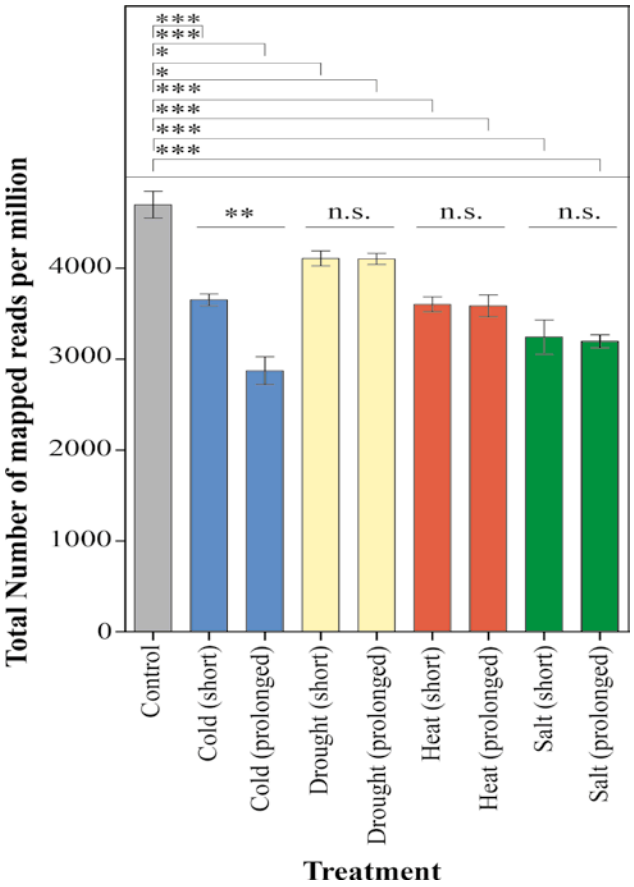

Fig. 1. Total number of mapped reads (per million of reads) on the 828 full-length LTR-REs of *P. trichocarpa*, in leaves of control and stress-exposed plants. Stresses included cold, drought, heat and salt treatments, for short and prolonged times. The differences between control and each treatment and between short and prolonged stress treatments were significant at $p<0.001$ (***), $p<0.01$ (**), $p<0.05$ (*), or not significant (n.s.) according to Tukey's test.

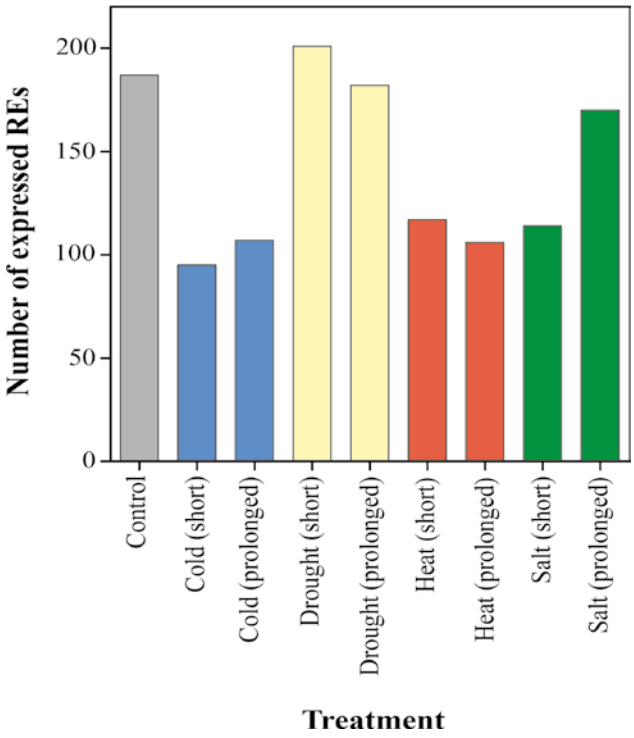

Fig. 2. Number of expressed (MRpM > 1) LTR-REs in leaves of control and stress-exposed plants. Stresses included cold, drought, heat and salt treatments, for short and prolonged times.

number of expressed LTR-REs increased after prolonged salt treatment.

The relationship between the abundance of a retrotransposon in the genome and its expression

In another analysis we measured the relationship between the abundance of a LTR-RE in the genome and the corresponding expression level. Such data are reported for control leaves in Fig. 3. It can be observed that abundant LTR-REs (average coverage > 100) are lowly expressed. Similar results were also found in leaves of plants exposed to different stresses (not reported).

Since assessing the expression of a LTR-RE is based on the occurrence of LTR-RE sequences in cDNA libraries, one might ascribe such occurrence to genomic DNA contamination. Actually, since the most abundant LTR-REs resulted slightly or even not expressed, contamination by genomic DNA in the cDNA libraries can be largely ruled out.

Influence of chromosomal localization on retrotransposons expression

In order to verify whether active LTR-REs were localized at specific chromosomal sites, we aligned LTR-RE sequences, highly expressed (MRpM > 10) in the control leaves, to the genome of *P. trichocarpa* (Fig. 4). For comparison, we separately aligned the genome with all the 828 LTR-REs; furthermore we determined the putative position of the centromere on each linkage group (LG) by aligning a tandemly repeated centromeric sequence of *P. trichocarpa* (Islam-Faridi et al. 2009) and *Chromovirus* elements (which preferentially localize at centromeres, Neumann et al. 2011). The chromosomal profiles of all the LTR-REs and highly expressed LTR-REs were substantially similar (Fig. 4). In some cases, minor peaks in the general LTR-REs profiles were apparently absent in the expressed LTR-REs profiles, suggesting that elements at those loci were generally inactive.

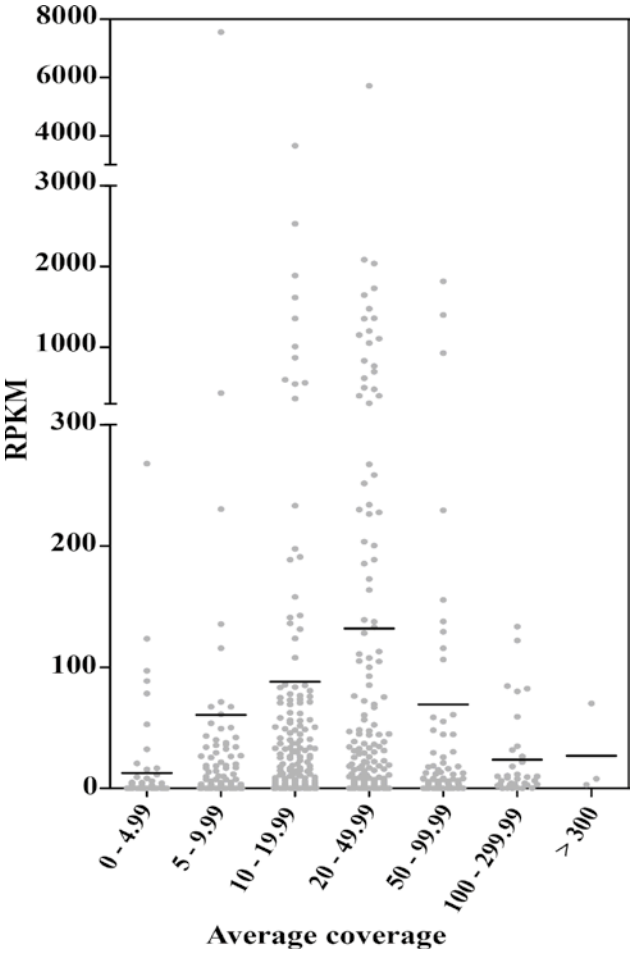

Fig. 3. Relationship between average coverage and RPKM for each of 828 full-length LTR-REs of *P. trichocarpa*.

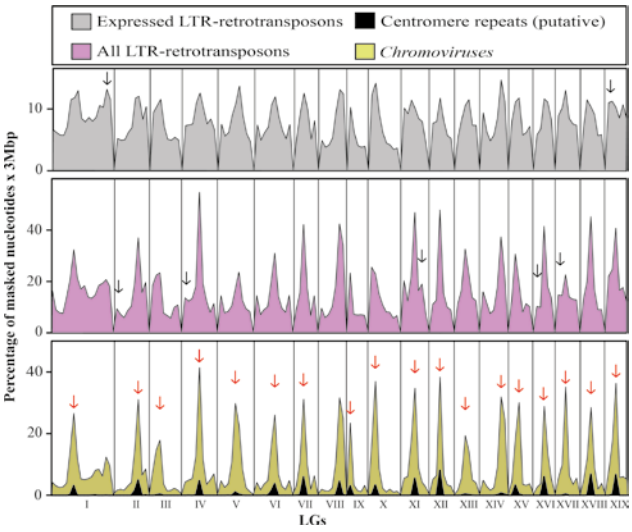

Fig. 4. Percentage of aligned nucleotides along *P. trichocarpa* LGs using all the full-length LTR-REs expressed in control leaves, all isolated full-length LTR-REs, all isolated *Chromovirus* REs and a putative centromeric sequence (in black). Red arrows indicate the putative position of the centromeres. Black arrows indicate minor peaks which are especially evident in the expressed LTR-REs profiles or in the profiles of all LTR-REs.

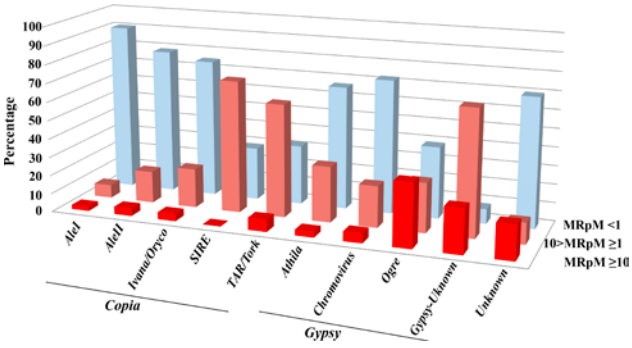

Fig. 5. Percentages of highly expressed (MRpM ≥ 10), expressed (MRpM ranging from 1 to 10) and unexpressed (MRpM < 1) LTR-REs, distinguished among LTR-RE superfamilies and lineages.

Only two peaks (within LG I and LG XIX) were apparently more evident in the expressed LTR-REs profiles, indicating that REs at these loci were particularly active. In general, it can be observed that peaks in putative centromere positions were less evident in the expressed LTR-RE profiles, suggesting that centromere LTR-REs were less active than elements lying at other loci (Fig. 4).

Influence of the superfamily/lineage of the retrotransposon on its expression

In order to assess whether the expression of LTR-REs was related to the superfamily/lineage to which the element belonged, LTR-REs were subdivided into lineages and separated among highly expressed (i.e., mapped by more than 10 reads per million), expressed (1-10 mapped reads per million) and not expressed (less than 1 mapped read per million). *Gypsy* elements resulted more expressed than *Copia*, in fact 48 out of 417 *Gypsy* REs (11.5%) were expressed, compared to 16 out of 368 *Copia* REs (4.4%). In general, most lineages showed low percentages of highly expressed or expressed LTR-REs (Fig. 5). However, for two *Copia* lineages (*SIRE* and *TAR/Tork*) and one *Gypsy* lineage (*Ogre*) the majority of LTR-REs resulted highly expressed or expressed (Fig. 5). In particular, the *Ogre* lineage showed the highest percentage of expressed elements. Diffused LTR-RE expression was also observed for those elements belonging to the *Gypsy* superfamily, but for which the lineage could not be determined.

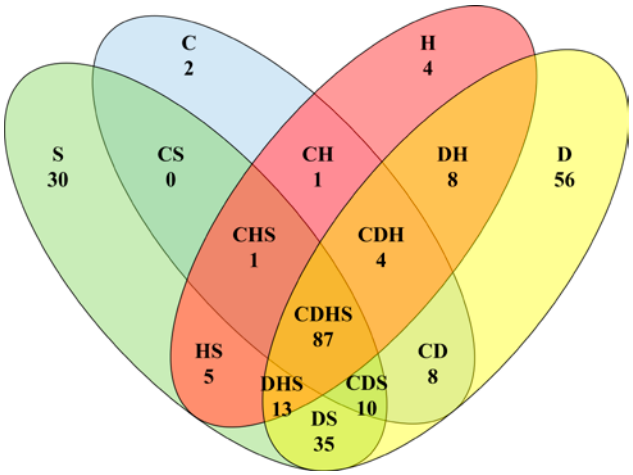

Fig. 6. Venn diagram of expressed (MRpM > 1) LTR-REs in the four stresses used in these experiments (D = drought; H = heat; C = cold; S = salt). Results of short and prolonged treatments were cumulated for each stress.

Stress-specific induction of retrotransposons expression

Most LTR-REs which were expressed in control leaves were also expressed in leaves of stress exposed plants. Of 313 LTR-REs expressed (MRpM > 1) in controls and/or in different stresses, only 4 (1.3%) were expressed only in controls and 81 (25.9%) were specifically activated by one or more stresses. Figure 6 reports the number of LTR-REs expressed (i.e. with more than one mapped read per million) after different stresses (in both short and prolonged treatments). Of 264 LTR-REs expressed during one or more stresses, 87 (33.0%) were active under each stress. Fifty-six elements (21.2%) were specifically active during drought treatments, 30 (11.4%) during salt treatments and 35 (13.3%) during both drought and salt stresses, indicating that these treatments were the most effective in inducing LTR-RE expression. On the contrary, cold and heat stresses induced only a limited number of LTR-REs (Fig. 6).

We also analysed differential expression of LTR-REs during the different stresses compared to the controls. Considering only the 72 highly expressed elements (MRpM > 10), 70 showed differential expression (fold change > 2, FDR-corrected $p < 0.05$) in at least one treatment (Fig. 7). No elements were differentially expressed along all treatments. In most cases, the same LTR-RE was under-expressed (blue in Fig. 7) or unaffected (white in Fig. 7) during the different stresses. Only 7 LTR-REs were induced (red in Fig. 7) or unaffected. Thirteen elements were repressed by certain treatments, activated by other, or unaffected (blue, red, or white in Fig. 7).

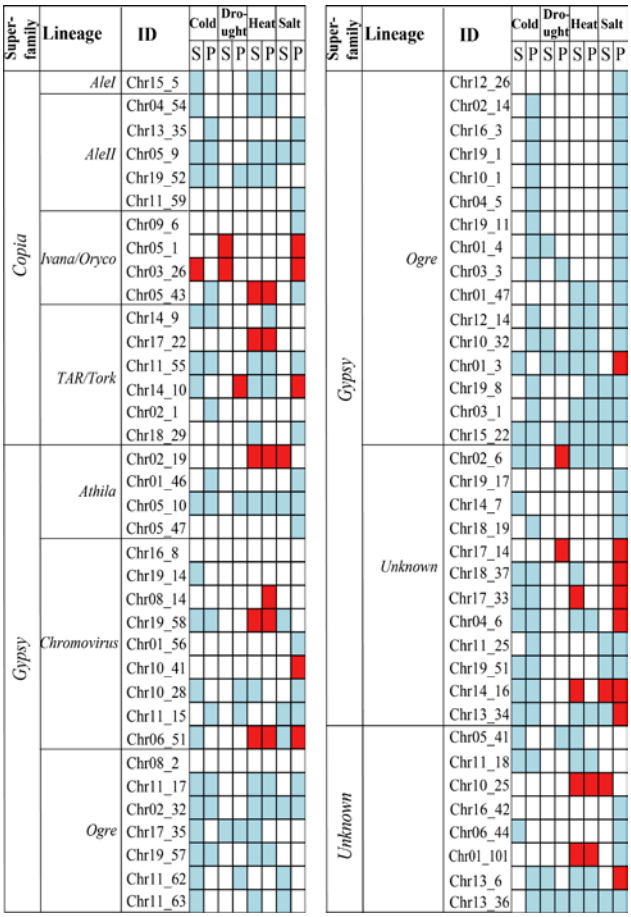

Fig. 7. Differential expression (fold change > 2, FDR-corrected $p < 0.05$) of LTR-REs after short (S) or prolonged (P) treatments with different stresses compared to controls. Blue cells refer to under-expression, red cells to over-expression, white cells indicate no effect of the treatment in comparison to control.

DISCUSSION

Availability of the updated sequence of the *P. trichocarpa* genome and of genomic DNA and cDNA libraries obtained from plants of the same genotype and subjected to different treatments, allowed us to evaluate several factors which can influence the expression of LTR-REs in this species.

In general, our data confirmed that the expression of retrotransposons is generally limited: only 72 out of 828 LTR-REs were mapped by more than ten reads per million. The transcription of REs have been reported in tissues and organs of many plant species (Grandbastien 2015), related to biotic and abiotic stresses or even without apparent induction. In rice, sunflower, *Citrus sinensis*, and even in poplars certain LTR-REs are actively transcribed (Rico-Cabanas and Martínez-Izquierdo

2007; Vukich et al. 2009; Gao et al. 2015; Giordani et al. 2016). However, the majority of LTR-REs are barely expressed (Vicient et al. 2001; Ishiguro et al. 2014; Vangelisti et al. 2019).

In some cases, specific LTR-RE sublineages have been shown to be activated and possibly over-expressed by different culture conditions, as tissue culture (Kashkush et al. 2003; Liu et al. 2004), wounding, methyl jasmonate and fungal elicitors (Takeda et al. 1999), various phytohormones and cold stress (He et al. 2010, 2012), heat stress (Ito et al. 2013). Hormones, and biotic/abiotic stresses induced a general LTR-RE activation in pine (Voronova et al. 2014; Fan et al. 2014) and in sunflower (Vangelisti et al. 2019). In the present study, as in all previous works, the same treatment up-regulated certain LTR-REs and repressed or unaffected other elements.

In *P. trichocarpa*, the overall RE expression level was higher in leaves of control plants than in those of plants exposed to different stresses, suggesting that plants responded to stresses increasing defence mechanisms related to REs. This is different from what found by Vangelisti et al (2019) in roots of sunflower: in this species, the expression level of LTR-REs remained substantially very low but it slightly increased after different stresses. Although the generally low level of LTR-REs expression, more than 40 elements showed a significant activity (more than 10 mapped reads x million), either in controls and stressed plants, suggesting that they are not silenced and hence may still have mutagenic potential, if retrotranscription and insertion in new sites would occur after expression.

The comparison, for each LTR-RE, of the abundance in the genome and its expression in leaves of control or stressed plants, showed that most expressed elements are generally lowly abundant. Such lack of correlation between LTR-RE abundance and transcription is not surprising: other studies showed that the more an element is repeated the more it is recognized by the RNA silencing machinery (Meyers et al. 2001; Yamazaki et al. 2001; Lisch 2009).

Low levels of transcription of repeated sequences are often attributed to DNA contamination of RNA samples. The low expression level of most abundant LTR-REs suggested also that the occurrence of retrotransposon-related reads in the cDNA libraries was not due to DNA contamination.

Genome localization of highly expressed (MRpM > 10) LTR-REs indicated that, in poplar, the expression of an element is only slightly related to its chromosomal localization, because the profiles of expressed LTR-REs parallels those of all LTR-REs. However, we observed a few specific chromosome regions showing differences between profiles of all the LTR-REs and expressed LTR-REs, suggesting that some regions are specifically activated or repressed. In species with much larger genomes than poplar, as the sunflower, LTR-RE expression was observed in specific genomic regions, relatively distant from putative centromeres, and preferentially located at chromosome ends (Mascagni et al. 2019).

Concerning the relationship between expression and superfamily/lineage of the elements, our results showed that expression of *Gypsy* REs was higher than expression of *Copia* elements. At lineage level, *Ogre* LTR-REs were by far the most transcribed elements. Among *Copia* lineages, the most expressed were *SIRE* and *TAR/Tork*, indicating that, besides chromosomal localization and genome abundance, also the "genotype" of the LTR-RE may play a role in its activation. Our results confirmed what previously shown in other studies, since many of the LTR-REs expressed in other species are actually of the *Copia* superfamily (Ma et al. 2008). In the case of tobacco, both *Tnt1* and *Tto1* (which are induced by tissue culture) belong to the *TAR/Tork* lineage (Neumann et al. 2019). *Gypsy* LTR-RE induction was reported in cotton (Hawkins et al. 2006), one of the families analyzed in that study belonged to the *Ogre* lineage. It can be concluded that, probably, different LTR-RE lineages are specifically activated in different species.

It can be assumed that young LTR-REs are more prone to be expressed than older elements, probably because the host needs time to develop defence mechanisms against new elements. *Ogre* and *TAR/Tork* elements are the youngest LTR-REs in *P. trichocarpa* (Mascagni et al. 2018b): this could explain why these two lineages showed the highest percentages of expressed elements.

Although most LTR-REs were expressed at the same level in plants subjected to different treatments, two stresses (salt and drought) specifically induced a number of LTR-REs. No elements were always induced or always repressed by every stress. In some cases, the same element was up-regulated by one stress and repressed by another, probably because of the occurrence, within the LTRs, of *cis*-regulatory motifs recognized in specific stresses, as those identified in the LTR of the *Ha*CRE1 element of sunflower (Buti et al. 2009).

In conclusion, this study outlines a general picture of LTR-RE activity in leaves of poplar plants treated with different stresses. Results allowed us to have a global insight on the features that affect LTR-RE expression. Since LTR-RE expression is just the first stage of retrotransposition, further studies are necessary to estimate subsequent stages of retrotransposition, including the

insertion of new elements in the genome, in order to clarify the biological significance of retrotransposon activity.

FUNDING

Research work funded by Department of Agriculture, Food, and Environment, University of Pisa, Italy, project PLANTOMICS.

DATA AVAILABILITY STATEMENT

The set of 828 full length LTR-REs of *P. trichocarpa* is available at the sequence repository of the Department of Agriculture, Food and Environment, University of Pisa (http://pgagl.agr.unipi.it/sequence-repository/).

REFERENCES

Baggerley KA, Deng L, Morris JS, Aldaz CM. 2003. Differential expression in SAGE: accounting for normal between-library variation. Bioinformatics. 19: 1477–1483.

Barghini E, Mascagni F, Natali L, Giordani T, Cavallini A. 2015a Analysis of the repetitive component and retrotransposon population in the genome of a marine angiosperm, *Posidonia oceanica* (L.) Delile. Mar Genomics. 24: 397–404.

Barghini E, Natali L, Giordani T, Cossu RM, Scalabrin S, Cattonaro F, Šimková H, Vrána J, Doležel J, Morgante M, Cavallini A. 2015b. LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. DNA Res. 22: 91-100.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statistic Soc. 57: 289-300.

Bennetzen JL. 2000. Transposable elements contributions to plant gene and genome evolution. Plant Mol Biol. 42: 251–269.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30: 2114-2120.

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P. 2012. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell. 24: 1242-1255.

Buti M, Giordani T, Vukich M, Gentzbittel L, Pistelli L, Cattonaro F, Morgante M, Cavallini A, Natali L. 2009. HACRE1, a recently inserted *Copia*-like retrotransposon of sunflower (*Helianthus annuus* L.). Genome. 52: 904-911.

Buti M, Moretto M, Barghini E, Mascagni F, Natali L, Brilli M, Lomsadze A, Sonego P, et al. 2018. The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry). GigaScience. 7: 1-14.

Chang W, Jääskeläinen M, Li S, Schulman AH. 2013. BARE retrotransposons are translated and replicated via distinct RNA pools. PLoS ONE. 8: e72270.

Cossu RM, Buti M, Giordani T, Natali L, Cavallini A. 2012. A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. Tree Genet Genomes. 8: 61–75.

Dieguez MJ, Vaucheret H, Paszkowski J, Mittelsten Scheid O. 1998. Cytosine methylation at CG and CNG sites is not a prerequisite for the initiation of transcriptional gene silencing in plants, but it is required for its maintenance. Mol Gen Genet. 259: 207–215.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics. 9: 18.

Falchi R, Vendramin E, Zanon L, Scalabrin S, Cipriani G, Verde I, et al. 2013. Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. Plant J. 76: 175-187.

Fan FH, Cui BW, Zhang T, Ding GJ, Wen XP. 2014. LTR-retrotransposon activation, IRAP marker development and its potential in genetic diversity assessment of masson pine (*Pinus massoniana*). Tree Genet Genomes. 10: 213–222.

Filichkin SA, Hamilton M, Dharmawardhana PD, Singh SK, Sullivan C, Ben-Hur A, Reddy ASN, Jaiswal P. 2018. Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching. Front Plant Sci. 9: 5.

Gao X, Zhou J, Li J, Zou X, Zhao J, Li Q, et al. 2015. Efficient generation of marker-free transgenic rice plants using an improved transposon-mediated transgene reintegration strategy. Plant Physiol. 167: 11-24.

Giordani T, Cossu RM, Mascagni F, Marroni F, Morgante M, Cavallini A, Natali L. 2016. Genome-wide analysis of LTR-retrotransposon expression in leaves of Populus × canadensis water-deprived plants. Tree Genet Genomes. 12: 75.

Grandbastien MA. 1998. Activation of plant retrotransposons under stress conditions. Trends Plant Sci. 3: 181–189.

Grandbastien MA. 2015. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. Biochim Biophys Acta. 1849: 403–416.

Hawkins JS, Kim HR, Nason JD, Wing RA, Wendel JF. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res. 16: 1252-1261.

He P, Ma Y, Dai HY, Li LG, Liu YX, Li H, Zhao GL, Zhang ZH. 2012. Characterization of the hormone and stress-induced expression of FaRE1 retrotransposon promoter in strawberry. J Plant Biol. 55: 1–7.

He P, Ma Y, Zhao G, Dai H, Li H, Chang L, Zhang ZH. 2010. FaRE1: a transcriptionally active *Ty1-copia* retrotransposon in strawberry. J Plant Res. 123: 707–714.

Ishiguro S, Ogasawara K, Fujino K, Sato Y, Kishima Y. 2014. Low temperature-responsive changes in the anther transcriptome's repeat sequences are indicative of stress sensitivity and pollen sterility in rice strains. Plant Physiol. 164: 671–682.

Islam-Faridi MN, Nelson CD, DiFazio SP, Gunter LE, Tuskan GA. 2009. Cytogenetic analysis of Populus trichocarpa - ribosomal DNA, telomere repeat sequence, and marker-selected BACs. Cytogenet Genome Res. 125: 74–80.

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.

Ito H. 2013. Small RNAs and regulation of transposons in plants. Genes Genet Syst. 88: 3–7.

Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet. 33: 102–106.

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

Lisch D. 2013. How important are transposons for plant evolution? Nat Rev Genet. 14: 49-61.

Liu Z, Han PF, Tan M, Shan XH, Dong YZ, Wang XZ, Fedak G, Hao S, Liu B. 2004. Activation of a rice endogenous retrotransposon *Tos17* in tissue culture is accompanied by cytosine demethylation and causes heritable alteration in methylation pattern of flanking genomic regions. Theor Appl Genet. 109: 200–209.

Lu X, Chen D, Shu D, Zhang Z, Wang W, Klukas C, Chen L, Fan Y, Chen M, Zhang C. 2013. The differential transcription network between embryo and endosperm in the early developing maize seed. Plant Physiol. 162: 440–455.

Ma Y, Sun HY, Zhao GL, Dai HY, Gao XY, Li H, Zhang ZH. 2008. Isolation and characterization of genomic retrotransposon sequences from octoploid strawberry (*Fragaria* × *ananassa* Duch.). Plant Cell Rep. 27: 499–507.

Mascagni F, Cavallini A, Giordani T, Natali L. 2017. Different histories of two highly variable LTR retrotransposons in sunflower species. Gene. 634: 5-14.

Mascagni F, Vangelisti A, Giordani T, Cavallini A, Natali L. 2018a. Specific LTR-retrotransposons show copy number variations between wild and cultivated sunflowers. Genes. 9: 433.

Mascagni F, Usai G, Natali L, Cavallini A, Giordani T. 2018b. A comparison of methods for LTR-retrotransposon insertion time profiling in the *Populus trichocarpa* genome. Caryologia. 71: 85-92.

Mascagni F, Vangelisti A, Usai G, Giordani T, Cavallini A, Natali L. 2019. A computational genome-wide analysis of long terminal repeats retrotransposon expression in sunflower roots (*Helianthus annuus* L.). Submitted.

Meignin C, Bailly JL, Arnaud F, Dastugue B, Vaury C. 2003. The 5' untranslated region and gag product of Idefix, a long terminal repeat retrotransposon from *Drosophila melanogaster*, act together to initiate a switch between translated and untranslated states of the genomic mRNA. Mol Cell Biol. 23: 8246–8254.

Meyers BC, Tingey SV, Morgante M. 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res. 11: 1660–1676.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nature Meth. 5: 621–628.

Natali L, Cossu RM, Mascagni F, Giordani T, Cavallini A. 2015. A survey of *Gypsy* and *Copia* LTR-retrotransposon superfamilies and lineages and their distinct dynamics in the *Populus trichocarpa* (L.) genome. Tree Genet. Genomes. 11: 107.

Neumann P, Navrátilová A, Koblížková A, Kejnovský E, Hřibová E, Hobza R, Widmer A, Doležel J, Macas J. 2011. Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA. 2: 4.

Neumann P, Novák P, Hoštáková N, Macas J. 2019. Systematic survey of plant LTR retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob DNA. 10: 1.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a galaxy based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics. 29: 792–793.

Rico-Cabanas L, Martínez-Izquierdo JA. 2007. CIRE1, a novel transcriptionally active *Ty1-copia* retrotransposon from *Citrus sinensis*. Mol Genet Genom. 277: 365.

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. Nat Genet. 20: 43–45.

Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF et al. 2012. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree Populus trichocarpa. New Phytol. 196: 713–725.

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

Sonnhammer EL, Durbin R. 1995. A dot-matriprogram with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 167: GC1–GC10.

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, et al. 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. 5: e1000734.

Takeda S, Sugimoto K, Otsuki H, Hirochika H. 1999. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J. 18: 383–393.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 313: 1596-1604.

Usai G, Mascagni F, Natali L, Giordani T, Cavallini A. 2017. Comparative genome-wide analysis of repetitive DNA in the genus *Populus* L. Tree Genet Genomes. 13: 96.

Vangelisti A, Mascagni F, Giordani T, Sbrana C, Turrini A, Cavallini A, Giovannetti M, Natali L. 2019. Arbuscular mycorrhizal fungi induce the expression of specific retrotransposons in roots of sunflower (*Helianthus annuus* L.). PLoS ONE. 14: e0212371.

Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell. 11: 1769–1784.

Vicient CM, Jaaskelainen MJ, Kalendar R, Schulman AH. 2001. Active retrotransposons are a common feature of grass genomes. Plant Physiol. 125: 1283–1292.

Vitte C, Fustier MA, Alix K, Tenaillon MI. 2014. The bright side of transposons in crop evolution. Briefings Funct Genom. 13: 276-295.

Voronova A, Belevich V, Rungis D, Jansons A. 2014. Stress-induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris* L.) genome. Tree Genet Genomes. 10: 937–951.

Vukich M, Giordani T, Natali L, Cavallini A. 2009. *Copia* and *Gypsy* retrotransposons activity in sunflower (*Helianthus annuus* L.). BMC Plant Biol. 9: 150.

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, Schulman AH. 2007. A unified classification systemfor eukaryotic transposable elements. Nature Rev Genet. 8: 973–982.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35: W265–W268.

Yamazaki M, Tsugawa H, Miyao A, Yano M, Wu J, Yamamoto S, et al. 2001. The rice retrotransposon *Tos17* prefers low-copy-number sequences as integration targets. Mol Genet Genomics. 265: 336–344.

Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. New Phytol. 214: 1338-1354.