Interspecific hybridisation and LTR-retrotransposon mobilisation-related structural variation in plants: a case study

Gabriele Usai[a, 1], Flavia Mascagni[a, 1], Alberto Vangelisti[a], Tommaso Giordani[a], Marilena Ceccarelli[b], Andrea Cavallini[a,*], Lucia Natali[a,*]

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

[b] *Department of Chemistry, Biology and Biotechnology, University of Perugia, Via Elce di sotto 8, 06123 Perugia, Italy*

* Corresponding authors at: Department of Agriculture, Food and Environment, University of Pisa, Pisa, Italy

E-mail addresses: andrea.cavallini@unipi.it (A. Cavallini); lucia.natali@unipi.it (L. Natali).

[1] These authors contributed equally to this work

ABSTRACT

The dynamics of long-terminal-repeat retrotransposons in two poplar species (*Populus deltoides* and *P. nigra*) and in an interspecific hybrid, recently synthesised, were investigated by analysing the genomic abundance and transcription levels of a collection of 828 full-length retroelements identified in the genome sequence of *P. trichocarpa*, all occurring also in the genomes of *P. deltoides* and *P. nigra*. Overall, genomic abundance and transcription levels of many retrotransposons in the hybrid resulted higher or lower than expected by calculating the mean of the parental values. A bioinformatics procedure was established to ascertain the occurrence of the same retrotransposon loci in the three genotypes. The results indicated that retrotransposon abundance variations between the hybrid and the mean value of the parents were due to i) co-segregation of retrotransposon high- or low-abundant haplotypes; ii) new retroelement insertions; iii) retrotransposon loss. Concerning retrotransposon expression, this was generally low, with only 14/828 elements over- or under-expressed in the hybrid than expected by calculating the mean of the parents. It is concluded that interspecific hybridisation between the two poplar species determine quantitative variation and differential expression of some retrotransposons, with possible consequences for the genetic differentiation of the hybrid.

Keywords:

Poplar

Interspecific hybrid

Long-terminal-repeats retrotransposons

Retrotransposon transcription

Retrotransposon-related structural variation

Retrotransposon insertions

Abbreviations

RE	Retrotransposon
LTR-RE	Long terminal repeat-retrotransposon
RPKM	Reads Per Kilobase per Million mapped reads

1. Introduction

Transposons are mobile genetic elements that can change their position in chromosomes through a process called transposition, in which each step is operated by specific enzymes encoded by transposons. In this sense, transposons can be distinguished into autonomous transposons, which encode the proper enzymes, and non-autonomous ones, which do not encode functional enzymes, but can transpose using enzymes produced by other elements [1Wicker et al 2007]. Depending on the transpositional mechanism, transposons are also distinguished into DNA transposons and retrotransposons (REs). The first transpose by excision of the element from its locus and insertion into another locus; REs use a replicative mechanism involving an RNA intermediate, which is retrotranscribed to cDNA and inserted into a new locus, leaving the original copy in its chromosomal site [2Bennetzen 2000]. This replicative mechanism determines an increase in RE copy number and, together with processes of whole-genome duplication, it is the main process responsible for the huge genome size increase which has often accompanied the evolution of most plant species [3SanMiguel et al 1998, 4Vicient et al. 1999].

In plants, the most abundant REs belong to the order of long terminal repeat (LTR) retrotransposons [1Wicker et al. 2007]. These elements possess two long, direct repeats at their ends, which flank a central portion containing sequences encoding the enzymes necessary for retrotransposition.

Plant LTR-REs are divided into two superfamilies, *Gypsy* and *Copia* [1Wicker et al. 2007], based on the order of protein coding domains within the pol gene. Superfamilies are in turn classified into different major lineages, depending on the sequence similarity of the coding regions [5-8Wicker and Keller 2007, Llorens et al 2011, Mascagni et al. 2017a,b]. In Angiosperms, the main *Gypsy* lineages are *Chromovirus*, a lineage of REs carrying a chromodomain at the 5' end of the coding portion, which is especially abundant in centromeres [6, 9Gorinsek et al. 2004; Llorens et al. 2011]; *Athila*, reported also in Gymnosperms [10 Neumann et al. 2019); and *Ogre*, represented by large elements with an open reading frame located upstream of the gag domain [11Neumann et al. 2003). The most diffused lineages of *Copia* superfamily are *Ale* (on its turn often distinguished into *AleI*/*Retrofit*/*Hopscotch* and *AleII*), *Ivana*, *Angela*, *Bianca*, *TAR* and *Tork* (often considered as an unique lineage), and *SIRE* [5, 10]Wicker and Keller 2007; Neumann et al. 2019).

The DNA sequence similarity within a lineage is, however, minimal and limited to the coding regions [1Wicker et al 2007]. In case the majority of the length of two elements shows high

sequence similarity (i.e. similarity is not limited to the coding portion), such elements belong to one and the same family, according to the rules proposed by Wicker et al. [1Wicker et al 2007].

In autonomous LTR-REs, the coding portion includes the gag and the pol domains, the former encoding virus-like particles, the latter a retrotranscriptase (RT), an RNaseH, a protease and an integrase, necessary to produce a double-stranded DNA and to integrate such DNA into the genome of the host; various *cis* features, including transcriptional promoter elements in the 5'-LTR, are necessary for LTR-RE transposition [2Bennetzen 2000].

The replicative mechanism of LTR-REs starts with the transcription of the element by RNA polymerase II, driven by a promoter in the 5'-LTR. Transcription of REs has been reported in a number of plant species, especially after exposure to various stresses [12Grandbastien 2015]. However, RE expression is generally much lower than that of functional genes [13-15Wessler 1996, Jaaskelainen et al 1999, Giordani et al 2016], mainly because of transcriptional or post-transcriptional repression, possibly related to chromatin methylation and to inactivation by RNA interference [16, 17Okamoto and Hirochika 2001, Lisch 2009]. In most plant species, LTR-REs are weakly expressed in standard culture conditions [18-21 Meyers et al 2001, Vicient et al 2001, Ishiguro et al 2014, Vangelisti et al 2019]. Generally, LTR-RE expression studies refer to specific elements [19, 22-25 Vicient et al 2001, Rico-Cabanas and Mart.Izq 2007, Ramallo et al 2008, Buti et al 2009, Kawakami et al 2011] or are limited to the identification of transposable sequences in the transcriptomes [26, 27 Parchman et al 2010, Lu et al 2013], while few studies have focused on genome-wide analyses of LTR-RE expression [15, 18, 28, 29 Giordani et al 2016, Meyers et al 2001, Marcon et al 2015, Jiang et al 2016].

Retrotransposon transcription is only the first step in the mobilisation of an RE. Such mobilisation is accomplished only when the mRNA has been reverse-transcribed to DNA and then reinserted into the genome, causing a permanent variation of the DNA. Despite widespread RE expression, new insertions in the genome (i.e. not accumulated during long evolutionary time spans) have been described in a few cases. For example, tobacco *Tnt1* and *Tto1* and rice *Tos17* mobilisation induced by tissue culture has been described [30, 31 Hirochika 1993, Grandbastien 1998]. The mobilisation of a *Copia* LTR-RE has been shown in sunflower, apparently not induced by stresses or particular culture conditions [32Vukich et al 2009].

In eukaryotic genomes, RE copy number variations are frequent, even among individuals of the same species [33, 34 Mascagni et al 2015, 2018]. These variations, strictly related to species evolution, do not include only RE accumulation, but are often counterbalanced by RE loss [35

Wang and Dooner 2006]. RE loss might be caused by unequal homologous recombination between the two LTRs of an element, a process which generally produces the so-called “solo-LTRs” [36 Vitte and Panaud 2003]. Unspecific DNA loss (including LTR-RE sequences) can be determined by illegitimate recombination [37 Devos et al 2002].

Variations in RE copy numbers within a species may also derive by combination of haplotypes presenting different numbers of inserted elements [7, 34, 38, 39 Mascagni et al 2017a, 2018, Brunner et al 2005, He et al 2006]: for example, if parents were heterozygous for LTR-RE insertions, i.e. they have haplotypes with different numbers of LTR-REs, then, when two haplotypes with many LTR-REs combine, the hybrid will have a greater number of retrotransposons than expected by calculating the mean of the parents. Such increases in copy number do not imply activation of REs and can be accomplished even in a single generation.

When genetic materials of two species are combined, as in interspecific crosses, either gene inactivation or even elimination of DNA of one of the parental species may occur, as in the interspecific hybrid *Hordeum vulgare* x *H. bulbosum*, in which a progressive loss of *H. bulbosum* chromosomes has been described during hybrid embryo development [40 Kasha and Kao 1970].

Many plant species are allopolyploid, originated after an interspecific hybridisation [41 Lewis 1979]. Cases are known of allopolyploids showing a reduced genome size when compared to the expected multiple of the diploid ancestors [42, 43 Leitch and Bennett 2004, Parisod et al 2010] or other in which extensive changes in the number of members of certain gene families have occurred [44 Rieseberg et al 1995]. Such genome changes have certainly established during the millennia of evolution of plants. However, genome changes have been also observed during the early generations of some interspecific hybrids [43 Parisod et al 2010].

McClintock [45] showed that a genome may react to conditions for which it is unprepared, the so called "genomic shock", to which it responds in unexpected manner, for example through the activation of transposable elements. Among genomic shocks, she suggested that species crosses may be a potent source of genomic modification [45 McClintock]. It is known that LTR-REs are activated in certain plant species after interspecific hybridisation and polyploidy [43Parisod et al 2010]. Genomes can undergo structural changes early after their formation, and transposon mobilisation is involved in genome reorganisation [46-48 Doyle et al 2008, Freeling et al. 2012, Parisod and Senerchia 2012].

The occurrence of RE-related structural genomic changes in the early generations of an interspecific hybrid has been studied especially using transposon-display techniques and other

PCR-based molecular markers [49-54 Petit et al. 2010, Kashkush et al 2003, Wang et al 2005, Zou et al 2011, Paz et al 2015, Senerchia et al 2015]. High-throughput 454 sequencing has been used to study RE-dynamics in *Nicotiana sylvestris*, *N. tomentosiformis* and in the allopolyploid *N. tabacum* [55 Renny-Byfield et al 2011]. All these studies refer to annual, herbaceous species. Less known are the consequences of interspecific crosses between perennial species.

Interspecific hybrids of *Populus* species are known for their superior growth [56 Dillen et al 2009]. Highly heterotic *Populus* x *canadensis* plants have been selected and cloned after crossing two poplar species, *P. deltoides* and *P. nigra*, and are largely cultivated in Europe and North America [57 Monclus et al 2006].

We used one of these interspecific hybrids as a case study to evaluate the genetic variation related to RE mobilisation, which is established after hybridisation and the early generations of vegetative propagation of hybrids. Clones of a hybrid of *Populus* × *canadensis*, obtained around 20 years ago from the two parental trees (*P. deltoides* L155-079 × *P. nigra* 71077-308), were provided by INRA, Orleans (France). We applied massive parallel sequencing and bioinformatics procedures to study the dynamics of a large set of poplar full-length LTR-REs during interspecific hybridisation. As a reference, we prepared and used a library of full-length LTR-REs of *P. trichocarpa*, which is phylogenetically closely related to both *P. deltoides* and *P. nigra*, having diverged from them only 8–13 million years ago [58 Sterck et al 2005]. Our study provided a first insight into the mechanisms by which LTR-REs change their abundance during interspecific hybridisation and subsequent vegetative propagation.

2. Materials and Methods

2.1. Collection of full-length LTR-REs from the genome sequence of Populus trichocarpa

Putative full-length LTR-REs were identified in the version GCA_000002775.3 [59 Zeng et al 2017] of the sequenced genome of *P. trichocarpa* [60, 61 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), using two procedures. The first consisted of the use of LTRharvest [62 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. In other analyses, the LTR-FINDER software [63 Xu and Wang 2007] was used, under default parameters, using a tRNA sequence collection of *P. trichocarpa*.

Around 20% of putative LTR-REs were randomly chosen and manually validated using DOTTER [64 Sonnhammer and Durbin 1995], verifying the occurrence of LTRs, dinucleotides TG and CA at the respective 5' and 3' ends, and TSDs. LTR-REs were annotated by BLASTN [65 Zhang et al. 2000) search against previously published plant RE datasets [66-69 Usai et al 2017, Barghini et al 2015, Natali et al 2015, Buti et al 2018] and using the Domain Search tool of RepeatExplorer [70 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.

2.2. Plant material and isolation of genomic DNA and RNA

Rooted cuttings of *Populus deltoides* (female parent, id. code L155-079), *P. nigra* (male parent, id. code 71077-308) and one hybrid of theirs (*P. x canadensis*, id. code DxN661200585, hereafter called PxC1, produced in 1998), kindly provided by INRA, Orleans (France), were cultivated in 20 × 20-cm^2 pots in the greenhouse under natural daylight conditions (750 µm m^{-2} s^{-1}, maximal photon flux density), with air temperature maintained at 17-29°C and relative humidity from 55 to 90%.

Fully expanded leaves, six to eight internodes from the apex, were collected from normally watered plants, 50-70 cm in height, at the same time of day (11.00 a.m.) and used for DNA (a single plant per genotype) and RNA isolation (three plants per genotype). Genomic DNA was extracted as described by Doyle and Doyle [71 1989].

Total RNA was isolated from leaves, according to the method described by Logemann et al. [72 1987], followed by DNase I (Roche) treatments according to the manufacturer's instructions to completely remove genomic DNA contamination. The RNA was then purified by following standard procedures [15 Giordani et al 2016].

2.3. Estimation of the genome size of the hybrid and its parents

Leaves were collected from two plants for each parent and three plants for the hybrid and fixed in ethanol: acetic acid (3:1, v/v). Leaf fragments were washed in an aqueous solution of 6 mM sodium citrate plus 4 mM citrid acid, treated with a mixture of 10% pectinase (Sigma) and 7%

cellulase (Calbiochem) in citrate buffer pH 4.6 for 60-90 min at 37°C and squashed under a coverslip in a drop of 60% acetic acid. The coverslips were removed after freezing at -80°C, and the air-dried preparations were concurrently Feulgen-stained after hydrolysis in 1 N HCl at 60°C for 8 min. Subsequently, the slides were subjected to three 10-min washes in SO_2 water prior to dehydration and mounting in distyrene-dibutylphthalatexylene (DPX; BDH Chemicals). Feulgen DNA absorptions in interphase nuclei were measured in images captured by a charge-coupled-device camera on a Leica DMRB microscope, using a Leica Q500MC image analyser.

2.4. gDNA sequence collection

Illumina paired-end sequencing data of genomic DNA of the three genotypes were downloaded from the SRA archive (https://www.ncbi.nlm.nih.gov/sra). In particular, the following libraries were collected: SRR3211864 (*P. deltoides*, female parent, L155-079), SRR3045878 (*P. nigra*, male parent, 71077-308) and SRR3747541, SRR3747542 (*P. x canadensis*, PxC1 hybrid, 661200585).

Libraries of paired-end reads of at least 6-M reads, 100 nt in length, were selected. Trimming on raw reads was performed using two procedures, the first treating them as single-end reads (for analyses of RE abundance) and the second as paired-end reads (for analyses of insertion sites). All sets of reads were checked for read quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, Illumina adapters and low-quality regions were removed using Trimmomatic, v. 0.38 [73 Bolger et al 2014] with the following parameters: ILLUMINACLIP:2:30:10, SLIDINGWINDOW:4:20, HEADCROP:10 and MINLEN:90. Organellar sequences were removed from the sequence sets by mapping against a database of chloroplast and mitochondrial sequences of poplar species [66 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. At the end of trimming, for each genotype, the same number of reads was randomly collected.

2.5. Kinship verification between hybrid and parents

The parental relationship of *P. deltoides* (plant L155-079) and *P. nigra* (plant 71077-308) towards their hybrid (PxC1, plant 661200585) was verified analyzing the genetic inheritance at gene level. Ten *Populus* sequences belonging to single-copy genes [74 Cossu et al., 2012) were randomly chosen. The datasets (29 Mb each) of trimmed Illumina reads of the three genotypes were merged and mapped against these gene fragments, using CLC-BIO GenomicWorkbench with the following parameters (mismatch cost = 1, insertion cost = 1, deletion cost = 1, length fraction = 0.9, similarity fraction = 0.9). For each gene, the mapped reads were aligned and visualized using ClustalX, v.2.0 [75 Larkin et al., 2007) with default parameters.

2.6. Illumina sequencing

The RNA-seq specific libraries were obtained as described previously [76 Cossu et al 2014]. For each parental and hybrid genotype, two biological replicates were collected with overall six cDNA libraries. The RNA-Seq libraries were produced using the TruSeq RNASeq Sample Prep kit according to the manufacturer´s protocol (Illumina Inc., San Diego, CA), as reported in a previous study [76Cossu et al 2014]. Single-end reads of 50 nt sequences were collected, and adapter sequences and low-quality bases were removed by using Trimmomatic with default parameters. Reads are deposited under PRJNA552799 bioproject accession.

2.7. Estimation of retrotransposon abundance

The abundances of specific LTR-REs in the genome of the three genotypes were analysed by aligning DNA reads (treating them as single-end reads, each sample made by the same number of reads, corresponding to 6.5 genome equivalents) of each genotype to the library of *P. trichocarpa* full-length LTR-REs. Alignment to the library was performed 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.

Differences in abundance among LTR-RE lineages were evaluated using Bonferroni's multiple comparison test.

2.8. Estimation of retrotransposon expression

The expression of LTR-REs was measured aligning cDNA sequence reads of the three genotypes to the library of *P. trichocarpa* full-length LTR-REs, using CLC-BIO Genomic Workbench 9.5.3 with the following parameters: mismatch cost 1, deletion cost 1, insertion cost 1, similarity 0.9 and length fraction 0.9. Non-uniquely mapping reads (i.e. reads that align with an equally good score at more than one sequence in the library) were assigned randomly to one of these sequences by the software. The expression level of each sequence was calculated and expressed both as mapped reads per million and RPKM [77 Mortazavi et al 2008]. Only LTR-REs, mapped by at least one read per million of reads in at least one sample, were considered as expressed [27 Lu et al 2013].

Expression values were compared considering RPKM values in the three genotypes using Baggerley's test [78 Baggerley et al 2003]. The weighted proportion fold changes between genotypes were considered significant when the weight of a sample was at least two-fold higher or lower than another, according to Baggerley's test, with a Bonferroni corrected p value ≤ 0.05. Such conservative parameters allowed us to establish significant differences between hybrid and parent plants, despite the small sample size (two replicates).

Differences in expression levels among LTR-RE lineages were evaluated using Bonferroni's multiple comparison test.

2.9. Computational identification of retrotransposon insertion sites

The pipeline for identifying retrotransposon insertion sites in the two parents and their hybrid is described in Figure 1a. For each analysed LTR-RE, Illumina 90-nt long reads of the three genotypes were aligned to the sequence of 27 nucleotides at the 5'-end of the 5'-LTR (including the dinucleotide TG, i.e. the first 27 nt of 5' LTR), using CLC-BIO Genomic Workbench with the following parameters: mismatch cost 1, insertion cost 1, deletion cost 1, length fraction 0.25, similarity fraction 0.95. All aligned reads were collected and aligned again to the full-length element (using CLC-BIO with the same parameters as above), and those mapping on 3'-LTR were discarded in order to reduce the complexity of the analysis. The remaining reads were in turn aligned to the 27-nt long fragment at the 5'-end of the 5'-LTR, and the read portions of each protruding at 5'-end were collected: these sequence fragments corresponded to the genomic DNA flanking the LTR-RE at its 5'-end, i.e. identified the site in which the LTR-RE is inserted. Aligning fragments (of the three genotypes) from each putative insertion site for each LTR-RE allowed us to

obtain seven possible patterns of alignment, which facilitated the calculation of the putative number of insertion sites of that element in the three genotypes. For each insertion site of an LTR-RE (Fig. 1a), it was assessed whether sequences of one, two or three genotypes were aligned.

The putative insertion sites in the hybrid were further checked adding each genomic fragment to the 27-nt long 5'-end of the LTR and mapping on these reconstructed sequences all Illumina reads of the parents (with CLC-BIO and using the same parameters as above): in case of a new insertion, no read of the parents should match these reconstructed sequences (Fig. 1b).

2.10. Experimental validation of retrotransposon new insertion sites

After establishing the insertion sites of an LTR-RE in the parents and in the hybrid, we used PCR to validate the presence or absence of such insertion sites in the three genotypes (Fig. 1c). For this, we randomly selected 9 putative LTR-RE insertion sites and designed "test" primers in the RE 5' flanking sequence (outside the RE, "test" forward primer) and in the 5'-LTR of that element ("test" reverse primer); PCR was carried out by using each primer pair on genomic DNAs of parents and hybrid. Briefly, 40 ng of genomic DNA were used in 20 µl PCR reaction with 2.5 mM MgCl2, 0.5 µM primers and 1 U Taq FirePol (Biodyne) DNA polymerase. Thermocycling was performed at 94°C for 30 s, at 58°C for 30 s and at 72°C for 10''.

Positive control PCR reactions were performed using as forward primer the reverse complement of the 5'-LTR "test" reverse primer and as reverse primer an oligonucleotide designed downstream of the forward primer in the LTR sequence (Fig. 1c).

Validation of a new insertion event in the hybrid genotype occurs when a PCR product with expected molecular weight is obtained when using "test" primer pairs in hybrid DNA, while in parental genomic DNAs, no PCR product is obtained when using "test" primer pairs, although a PCR product occurs in positive control PCR. The list of primers is available as Supplementary material # 1.

3. Results

3.1. Isolation and annotation of full-length LTR-REs of P. trichocarpa

The full-length LTR-REs used in this study were isolated from the genome sequence of *Populus trichocarpa*. For this species, a well-established and nearly complete genome sequence is available compared to other poplar species [59, 60Zeng et al 2017, Tuskan et al 2006]. *Populus trichocarpa* LTR-REs have already been shown to be present in the genomes of the species used in this study, *P. deltoides* and *P. nigra* [15Giordani et al 2016].

Our research group has already collected full-length retrotransposons from the *P. trichocarpa* genome, isolating 958 putative full-length elements using LTR-FINDER and DOTTER [68 Natali et al 2015]. However, during recent years, the genome sequence of *P. trichocarpa* has been updated, deciphering the sequences of a number of previously unresolved loci. Hence, in this work, we performed a completely new scan of the updated version of the poplar genome sequence to isolate full-length retrotransposons using LTRharvest and LTR-FINDER. In particular, LTRharvest was used with stringent parameters, including the occurrence of the dimers TG/CA at the 5' and 3' end of the putative LTR-RE. A sample of isolated elements (corresponding to 20%) were validated at the structural level using DOTTER. All these sequences resulted as LTR-REs.

The new dataset of full-length LTR-REs included 828 elements (Fig. 2). A multiFASTA 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/). The majority of LTR-REs belonged to the *Gypsy* superfamily (417/828), while *Copia* elements amounted to 368/828. For 43 full-length elements, it was not possible to identify the superfamily. The annotation procedure also allowed us to determine the specific lineages to which the isolated elements belonged. Although for a group of *Gypsy* LTR-REs the lineage could not be identified, we found elements belonging to three *Gypsy* lineages (*Athila*, *Ogre* and *Chromovirus*). Six *Copia* lineages (*Ale*, distinguished into *AleI* and *AleII*, *Angela*, *Bianca*, *Ivana*, *SIRE* and *TAR/Tork*) were identified.

3.2. Comparison of RE abundances between parents and hybrids

The three genotypes used in this study, *P. deltoides* (L155-079), *P. nigra* (71077-308) and hybrid PxC1 (DxN661200585), were first characterised at the cytological level. All plants showed a similar genome size (data not shown); therefore, the hybrid is to be considered diploid, as are the parents. The kinship between parents and hybrid was verified by comparing allelic Illumina sequence reads of the three genotypes after alignment to DNA fragments of ten randomly

selected single-copy genes of *Populus* [74 Cossu et al. 2012). Alignments are reported in Supplementary Materials # 2.

The extent of structural variations related to the mobilisation of LTR-REs, using high-throughput sequencing data, can be estimated comparing the increase or decrease of the coverage of a certain element in different genotypes [79 Alkan et al 2011]. We used three sets of 29 Mb 90-nt-long Illumina sequences of each of the three genotypes to map the set of full-length LTR-REs of *P. trichocarpa*. This procedure for estimating sequence abundance in a genome is commonly used in many species [7, 8, 33, 67, 80-82Mascagni et al 2017a, b, 2015, Barghini et al 2015, Swaminathan et al 2007, Tenaillon et al 2011, Barghini et al 2014], including poplar [68 Natali et al 2015].

The number of reads used for mapping corresponded to a 6.5 x coverage for each genotype. The average coverages (i.e. the sum up of the bases of the aligned part of all the reads divided by the length of the reference sequence) of 828 unique LTR-REs in the three genotypes are reported in Supplementary Materials # 3. The comparison between the average coverages of each LTR-RE between parents is reported in Figure 2. As expected, large differences were found in the abundances of LTR-REs between parents (r^2 = 0.63, Fig. 3). The LTR-RE average coverage distribution of the hybrid was also compared to that of the "expected" hybrid, calculated using the mean average coverage of parents (Fig. 3). The PxC1 hybrid showed a low correlation coefficient (0.46) because of the occurrence of numerous elements with lower abundance than expected by calculating the mean of the parents (Fig. 3).

The ratio between the LTR-RE average coverage of the hybrid and the mean average coverage of the parents is reported in Table 1, keeping separated the different lineages to which each LTR-RE belonged. Table 1 shows that some lineages are more prone than other to change their abundances in the hybrid in comparison to the expected calculating the mean between the parents, i.e. the lineages *Gypsy-Unknown*, *Copia-TAR/Tork*, *Copia-AleI*, *Copia-AleII*, *Gypsy-Chromovirus* and *Copia-Ivana*.

3.3. Identification of new RE insertion sites

As stated in the Introduction section, increase in the abundance of LTR-REs in a genome may be related to retrotransposition events and/or to the combination in a hybrid of different haplotypes. In the first case, it is to be assumed that retrotransposons have activated after

hybridisation; in the second case, the occurrence of large structural variations between homologous chromosomes in the parent(s), i.e., hemizygosity of LTR-RE insertion sites, is to be postulated. As a consequence of the latter assumption, when two haplotypes with many LTR-REs combine, the resulting hybrid will show a higher number of retrotransposons than expected based on the mean of the parents.

To establish if one or both these processes have occurred with the hybridisation, we performed a fine analysis of the insertion sites of 11 randomly selected LTR-REs, establishing a pipeline, described in Figure 1. The procedure consisted of retrieving the 27-nt-long fragment at the 5' end of the 5'-LTR of an element (including the dinucleotide TG, see Supplementary Materials # 4) and of collecting all Illumina 90-nt-long reads (of parents and hybrid) which aligned to it. These reads were aligned to the full-length element, and those mapping on 3'-LTR were discarded. The remaining reads were again aligned to the 27-nt-long fragment, and the portions of the reads not aligning at the 5' end were collected: these sequences correspond to the genomic DNA flanking the LTR-RE at its 5' end, i.e. the site in which the LTR-RE is inserted. These genomic sequences of the three genotypes were aligned, and the number of insertion sites of that element in the three genotypes was calculated. Five possible alignment patterns can be distinguished (Fig. 1a): 1) the site is represented by alignment of genomic sequences of all genotypes, meaning that an insertion site was present in both parents and was transmitted to their hybrid; 2) it is represented by sequences of the hybrid and of one or the other parent; in this case, it can be assumed that the insertion site existed only in one of the two parents and was transmitted to the hybrid; 3) the site is represented by alignment of sequences of both parents and not of the hybrid, indicating that the insertion site was hemizygous in both parents and therefore not transmitted to the hybrid or that both copies were lost in the hybrid; 4) the insertion locus was present only in one of two parents, indicating that the retrotransposon was hemizygous in that parent and not transmitted to the hybrid or that it was lost in the hybrid; 5) the insertion site was represented only in the hybrid, hence it was absent in the two parents and was the result of a new insertion event in the hybrid.

Putative new insertion sites in the hybrid were first validated producing a consensus sequence of the insertion site (including the dinucleotide TG and the 5'-end of the LTR-RE) and mapping on these sequences the two large coverage sets of reads of the parents. If reads from parental gDNAs were aligned to the genomic portion or to the LTR portion of the consensus

sequences but no single read was aligned to both the genomic and the LTR-RE portions of the consensus, the new insertion site was considered as validated (Fig. 1b).

The results of this analysis for 11 LTR-REs are reported in Table 2. The patterns from 1 to 3 represent sites occurring in at least one parent and regularly inherited by the hybrid. In total, the insertion sites of this kind accounted for 249 of the 536 analysed sites (46.5%). The patterns from 4 to 6 include sites for which the loss of the element or its hemizygosity in at least one of the parents should be postulated; these sites were the majority (278/536, 51.9%). Finally, in nine cases (1.7%), the insertion site was specific to the hybrid; hence, it was produced by a new insertion. Interestingly, seven out of nine new insertions are related to *Copia* LTR-REs (*TAR/Tork*, *Ivana*) and only two out of nine to *Gypsy* LTR-REs (of the *Chromovirus* lineage).

The occurrence of these nine new LTR-RE insertion sites was definitely validated by PCR, using "test" primers designed on the genomic portion (forward) and on the LTR-RE portion (reverse) of the consensus of the insertion locus (Fig. 1c). As a control, an LTR-RE oligonucleotide reverse complement to the previous primer (test reverse primer) was coupled to a primer (reverse) designed on the internal portion of the LTR-RE. In case of validation, with the first "test" primer pair, the amplification should have been obtained only in the hybrid, and, with the second "control" primer pair, it should be achieved also in the parents (Fig. 1c). All nine new insertion sites were validated using this procedure. For three of them, the electrophoretic patterns are reported in Figure 4.

3.4. Comparison of RE expression between parents and hybrids and relationship between RE abundance and expression

The expression of 828 poplar LTR-REs in *P. deltoides*, *P. nigra* and in their hybrid PxC1 was analysed by Illumina RNA-seq.

The LTR-RE expression was assessed by mapping reads onto LTR-RE sequences, using CLC-BIO Genomics Workbench. This tool randomly assigns non-specifically matched reads, i.e. those reads that align at more than one position with an equally good score. In our experiments, the average percentage of non-specific matches in the hybrid was 7.1 ± 1.4, and similar percentages were observed in the parents (data not shown). Hence, such non-specificity only slightly altered the results.

We measured the correlation between the RPKM values of each LTR-RE and the abundance in the hybrid and in the two parents (Suppl. Fig. 1). Considering the most abundant LTR-REs (i.e., with average coverage > 100), only one element in *P. nigra* and two elements in the hybrid resulted highly expressed. The other most abundant LTR-REs were not (or only slightly) expressed; correspondingly, the most expressed LTR-REs were poorly represented in the genomes of the hybrid. These data also indicate that contamination by genomic DNA in the cDNA libraries could be largely ruled out.

Figure 5 reports the RPKM values of 828 poplar LTR-REs in the two parents. The LTR-RE expression profiles were considerably different between the two parents, as shown by the correlation coefficient. Comparing the expression values of the hybrid vs. those obtained averaging the two values of parents, some LTR-REs were under- (7 LTR-REs) or over-expressed (7 LTR-REs) in the hybrid (Fig. 6), suggesting an expression rate higher or lower than the mean of the parents.

Three out of 7 over-expressed LTR-REs in the hybrid were considerably less abundant in the genome of the hybrid than expected considering the mean redundancy values between parents (see Suppl. Table 1). They include two *Gypsy* elements (of the *Athila* lineage) and one *Copia* RE (of the *Ivana* lineage).

The observed differences in expression levels among elements were not related to the RE lineage, because the ratios between RPKM in the hybrid and the mean RPKM between parents were not different among lineages (Suppl. Table 2).

4. Discussion

We investigated the behaviour of LTR-retrotransposons after interspecific hybridisation, an event that is supposed to produce a "genomic shock", as suggested by Barbara McClintock [45 1984]. As a matter of fact, the induction of DNA transposon mobilisation by interspecific hybridisation has already been hypothesized by McClintock based on her first studies on mobile elements. Less known are the effects of such genomic shock on the activity of retrotransposons, which represent the most abundant transposon class in plants, accounting for a large fraction of the genome. In particular, to our knowledge, studies on LTR-RE activation in interspecific hybrids of plants, using new high-throughput sequencing techniques, are still limited.

The occurrence of large variations in the retrotransposon contents and activities among and even within species has been largely ascertained in both annual and perennial species [33, 66,

83, 84 Mascagni et al 2015, Usai et al 2017, Neumann et al 2006, Piegu et al 2006] and even in poplar species [85 Pinosio et al 2016]. In some cases, such changes occurred after interspecific hybridisation [86 Ungerer et al 2006]. Usually, such variation is reported in the long evolutionary time scale. However, RE-related structural variations have been described also in newly produced hybrids.

In such synthetic hybrids and allopolyploids, created *de novo* mimicking natural species, the level of genome modification related to hybridization and allopolyploidy is apparently species-dependent. For example, several structural genome rearrangements were observed in allopolyploid *Tragopogon* hybrids [87, 88 Lim et al., 2008; Sarilar et al., 2013); in *Aegilops–Triticum* synthetic hybrids, elimination of DNA sequences from homologous chromosomes and gene loss have been commonly observed [89, 90 Shaked et al., 2001; Chantret et al 2005). In some hybrids of the genus *Spartina*, genome structural changes were found in the first few generations following interspecific hybridisation and allopolyploidy [91 Parisod et al 2009]. In newly synthesized *Brassica* hybrids, chromosomal repatterning have been reported [92, 93 Pires et al 2004, Udall et al 2005]. In some cotton interspecific hybrids, amplification or reduction of repetitive sequences was observed [49, 94, Petit et al 2010 Zhao et al 1998]. In young *Nicotiana* allopolyploids, transposable element activation occurred during the first generations of the allopolyploid [49, 95 Petit et al 2010, Mhiri et al. 2019). On the other hand, no significant structural variations were observed in newly produced allopolyploids in cotton [96 Liu et al., 2001), *Spartina* [97 Baumel et al. 2002), and wheat [98 Charles et al. 2008).

The hybrid analysed in our study is diploid and was obtained around 20 years ago, crossing *P. deltoides* with *P. nigra*. As poplar is an outcrossing species, parents are highly heterozygous, hence the hybrid can show large genetic differences compared to the parents due to segregation and recombination during sexual reproduction. The primary hybrid was then clonally propagated for some generations, and the same was done for the parental plants. Hence, the plants used in this study are individuals of the same age (3-year-old plants derived from cuttings), grown in the same conditions and belonging to three genotypes (two parents and one hybrid).

The analysed poplar hybrid showed huge differences in the abundance of certain LTR-REs in the hybrid compared to the expected values obtained calculating the mean of the parents. It is presumable that besides amplification or unequal homologous and illegitimate recombination, such differences are largely due to the segregation (in the hybrid) of parental haplotypes differently rich in such elements. As a matter of fact, our results are in agreement with the

occurrence, in the genomes of *P. deltoides* and *P. nigra*, of a number of hemizygous LTR-REs (i.e. present in a locus in a chromosome, but absent in the same locus in the homologous chromosome).

The occurrence in the genome of LTR-RE-related hemizygous structural variations has been assessed in different species, for example in maize [38 Brunner et al 2005]. The hemizygous condition of LTR-REs is apparently in relation to the reproduction mechanism of these elements, which insert randomly in one of the two homologous chromosomes. Most likely, LTR-RE hemizygosity is less frequent in autogamous plants (in which self-crossing favours the passage from the hemizygous to the homozygous or to the null condition), while it should be easily maintained across generations in allogamous species such as poplar.

Another process, which could have determined changes in the abundance of certain LTR-REs, might be the unequal recombination between LTRs of one and the same element [36 Vitte and Panaud 2003]. Our analyses did not allow us to distinguish, in a certain locus, between the presence of a full-length element and the presence of a solo-LTR.

Our results also showed that, besides segregation and recombination of LTR-REs, at least a part of the structural variations indicated by abundance variation of these elements is related to the production of new copies of LTR-REs, subsequent to interspecific hybridisation. Actually, in a sample of 11 randomly selected LTR-REs, we identified nine new insertion sites (all validated by PCR), indicating that LTR-RE mobilisation occurred frequently in the first clonal generations of the interspecific cross.

Most studies on the activation of REs after interspecific hybridization concern annual herbaceous species, which generally reproduce by sexual propagation. Sexual reproduction, passing through meiosis, might filter out many structural variations such as new insertions of REs, which should be hemizygous. The perennial habitus and the possibility of vegetative propagation, make the poplar a species potentially able to maintain a greater hemizygosity of the insertion sites compared to an annual species that reproduces every year. According to this hypothesis the hemizygosity level of the insertion sites should be higher in perennial than in annual species. This hypothesis will be tested by using new sequencing methods which produce very long sequences, allowing to decipher the two haplotypes of a diploid individual.

The three genotypes analysed in our study are not the original individuals that have been crossed and their hybrid, but clones of those three individuals, obtained by cuttings. It is not possible to exclude that the variations observed in the hybrid were at least in part induced also by

clonal propagation. However, it is reasonable that the activation of REs by clonal propagation is minimal. McClintock [45](1984) suggested that only *in vitro* culture propagation, which implies a huge reorganization of cells and tissues, could cause a genomic shock. On the other hand, while *in vitro* culture produces a large amount of phenotypic variants (the so-called somaclonal variation), "natural" propagation, for example by cuttings, is used with the aim to maintain the phenotype of the original plant, being the production of genetic variants with this type of propagation quite low. This seems all the more true in poplar, where *P.* x *canadensis* hybrids are regularly cultivated and are phenotypically very stable. If many REs were activated in clonal propagation, a certain number of phenotypic variants would be expected in the offsprings.

Concerning the different LTR-RE lineages involved in the variations, abundance variations (both through segregation and recombination and through new insertions) were at least in part related to the lineage (i.e. the genotype) of the element. For example, *Chromovirus* and *TAR/Tork* LTR-REs were more subject to changes than *Athila* LTR-REs. It is possible that elements belonging to certain lineages are more prone to be activated and, consequently, to be subjected to hemizygosity in parental species and/or to new insertions in the hybrid. It is known that young LTR-REs are more often active than old elements, probably because the host needs time to develop specific defence mechanisms. The *TAR/Tork* elements are the most recently active LTR-REs in *P. trichocarpa* [99 Mascagni et al 2018b?]. If this was true also in *P. deltoides* and *P. nigra*, it could explain why *TAR/Tork* elements showed large changes in abundance between hybrids and parents.

Overall, transcriptomics data showed a low expression level of LTR-REs, as often observed for these elements in many species [4, 21 Vicient et al 1999, Vangelisti et al 2019], including poplar [15Giordani et al 2016]. In the hybrid, as in the parents, the expression of an element was inversely related to its abundance. Such a lack of correlation between LTR-RE abundance and transcription is not surprising, in fact it is known that the more abundant is an element, the more easily it is subjected to RNA silencing (17, 18, 100 Meyers et al. 2001; Yamazaki et al. 2001; Lisch 2009).

Only a few elements (14 over 828) were over- or under-expressed in the hybrid, compared to the value obtained calculating the mean of the parents. In general, expression level was similar for every lineage, i.e. it did not depend on the lineage, but it was specific to the LTR-RE family. It can be hypothesized that the over- or under-expression of these elements is related to the new genomic asset of the hybrid and to the local epigenetic setting of each element. Concerning the

over-expressed elements, they did not show larger abundance in the genome of the hybrid than in the mean of the parents. Other cycles of clonal propagation would be necessary to verify if over-expression of these elements might produce new LTR-RE insertions or if their retrotransposition is blocked at post-transcriptional level.

As LTR-REs mobilisation depends on the preliminary transcription of the element, it can be hypothesised that the transcription of newly inserted elements occurred especially in the first years after the interspecific cross and/or in the first vegetative generation and is now almost completely ceased. However, the possibility that, in poplar hybrids, LTR-REs activity is still ongoing, even with low LTR-RE transcription rates, cannot be ruled out.

In conclusion, using a pipeline based on Illumina sequencing of genomic DNA, we showed the occurrence of structural variations related to LTR-RE mobilisation in the first clonal generations of a poplar interspecific hybrid. Studies are in progress to assess the gene contents of loci subjected to new insertions in order to evaluate the possible phenotype changes related to LTR-RE mobilisation.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This study was supported by the Department of Agriculture, Food and Environment, University of Pisa, Project “Plantomics”.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: XXXXXXXXXXXXXXX.

References

[1] T. Wicker, F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A.H. Schulman, A unified classification system for eukaryotic transposable elements, Nature Rev. Genet. 8 (2007) 973–982.

[2] J.L. Bennetzen, Transposable elements contributions to plant gene and genome evolution, Plant. Mol. Biol. 42 (2000) 251–269.

[3] P. SanMiguel, B.S. Gaut, A. Tikhonov, Y. Nakajima, J.L. Bennetzen, The paleontology of intergene retrotransposons of maize, Nature Genet. 20 (1998) 43–45.

[4] C.M. Vicient, A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, A. Beharav, E. Nevo, A.H. Schulman, Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*, Plant Cell 11 (1999) 1769–1784.

[5] T. Wicker, B. Keller, Genome-wide comparative analysis of *Copia* retrotransposons in *Triticeae*, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *Copia* families, Genome Res. 17 (2007) 1072–1081.

[6] C. Llorens, R. Futami, L. Covelli, L. Domínguez-Escribá, J.M. Viu, D. Tamarit, J. Aguilar-Rodríguez, M. Vicente-Ripolles, G. Fuster, G.P. Bernet, F. Maumus, A. Munoz-Pomer, J.M. Sempere, A. Latorre, A. Moya, The *Gypsy* database (GyDB) of mobile genetic elements: release 2.0, Nucl. Acids Res. 39 (2011) D70–D74.

[7] F. Mascagni, A. Cavallini, T. Giordani, L. Natali, Different histories of two highly variable LTR retrotransposons in sunflower species, Gene 634 (2017) 5–14.

[8] F. Mascagni, T. Giordani, M. Ceccarelli, A. Cavallini, L. Natali, Genome-wide analysis of LTR retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.), BMC Genomics 18 (2017) 634.

[9] B. Gorinšek, F. Gubenšek, D. Kordiš, Evolutionary genomics of chromoviruses in eukaryotes, Mol. Biol. Evol. 21 (2004) 781-798.

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

[11] P. Neumann, D. Pozarkova, J. Macas, Highly abundant pea LTR retrotransposon *Ogre* is constitutively transcribed and partially spliced, Plant Mol. Biol. 53 (2003) 399–410.

[12] M.A. Grandbastien, LTR retrotransposons, handy hitchhikers of plant regulation and stress response, Biochim. Biophys. Acta 1849 (2015) 403–416.

[13] S.R. Wessler, Plant retrotransposons: turned on by stress, Curr. Biol. 6 (1996) 3.

[14] M. Jääskeläinen, A.H. Mykkänen, T. Arna, C.M. Vicient, A. Suoniemi, R. Kalendar, H. Savilahti, A.H. Schulman, Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells, Plant J. 20 (1999) 413-422.

[15] T. Giordani, R.M. Cossu, F. Mascagni, F. Marroni, M. Morgante, A. Cavallini, L. Natali, Genome-wide analysis of LTR-retrotransposons expression in leaves of *Populus* x *canadensis* water-deprived plants, Tree Genet. Genomes 12 (2016) 75.

[16] H. Okamoto, H. Hirochika, Silencing of transposable elements in plants, Trends Plant Sci. 6 (2001) 527-534.

[17] D. Lisch, Epigenetic regulation of transposable elements in plants, Annu. Rev. Plant Biol. 60 (2009) 43‑66.

[18] B.C. Meyers, S.V. Tingey, M. Morgante, Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome, Genome Res. 11 (2001) 1660–1676.

[19] C.M. Vicient, M.J. Jaaskelainen, R. Kalendar, A.H. Schulman, Active retrotransposons are a common feature of grass genomes, Plant Physiol. 125 (2001) 1283–1292.

[20] S. Ishiguro, K. Ogasawara, K. Fujino, Y. Sato, Y. Kishima, 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 (2014) 671–682.

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

[22] L. Rico-Cabanas, J.A. Martinez-Izquierdo, CIRE1, a novel transcriptionally active *Ty1-copia* retrotransposon from *Citrus sinensis*, Mol. Gen. Genomics 277 (2007) 365–377.

[23] E. Ramallo, R. Kalendar, A.H. Schulman, J.A. Martínez-Izquierdo, *Reme1*, a *Copia* retrotransposon in melon, is transcriptionally induced by UV light, Plant Mol. Biol. 66 (2008) 137–150.

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

[25] T. Kawakami, P. Dhakal, A.N. Katterhenry, C.A. Heatherington, M.C. Ungerer, Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons, Genome Biol. Evol. 3 (2011) 156-157.

[26] T.L. Parchman, K.S. Geist, J.A. Grahnen, C.W. Benkman, C.A. Buerkle, Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery, BMC Genomics 11 (2010) 180.

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

[28] H.S. Marcon, D.S. Domingues, J. Costa Silva, R. Junqueira Borges, F. Filippi Matioli, M.R. de Mattos Fontes, C.L. Marino, Transcriptionally active LTR retrotransposons in *Eucalyptus* genus are differentially expressed and insertionally polymorphic, BMC Plant Biol. 15 (2015) 198.

[29] S. Jiang, D. Cai, Y. Sun, Y. Teng, Isolation and characterization of putative functional long terminal repeat retrotransposons in the *Pyrus* genome, Mob. DNA 7 (2016) 1.

[30] H. Hirochika, Activation of tobacco retrotransposons during tissue culture, EMBO J 12 (1993) 2521-2528.

[31] M.A. Grandbastien, Activation of plant retrotransposons under stress conditions, Trends Plant Sci. 3 (1998) 181–189.

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

[33] F. Mascagni, E. Barghini, T. Giordani, L.H. Rieseberg, A. Cavallini, L. Natali, Repetitive DNA and plant domestication: variation in copy number and proximity to genes of LTR-retrotransposons among wild and cultivated sunflower (*Helianthus annuus*) genotypes, Genome Biol. Evol. 7 (2015) 3368-3382.

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

[35] Q. Wang, H.K. Dooner, Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus, Proc. Natl. Acad. Sci. USA 103 (2006) 17644–17649.

[36] C. Vitte, O. Panaud, Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L., Mol. Biol. Evol. 20 (2003) 528–540.

[37] K.M. Devos, J. Brown, J.L. Bennetzen, Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*, Genome Res. 12 (2002) 1075–1079.

[38] S. Brunner, K. Fengler, M. Morgante, S. Tingey, A. Rafalski, Evolution of DNA sequence nonhomologies among maize inbreds, Plant Cell 17 (2005) 343–360.

[39] G. He, X. Luo, F. Tian, K. Li, Z. Zhu, W. Su, X. Qian, Y. Fu, X. Wang, C. Sun, J. Yang, Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice, Genome Res. 16 (2006) 618–626.

[40] K.J. Kasha, K.N. Kao, High-frequency haploid production in barley (*Hordeum vulgare* L.), Nature 225 (1970) 874-876.

[41] W.H. Lewis, Polyploidy in angiosperms: dicotyledons. In: W.H. Lewis (ed.) Polyploidy, biological relevance. Plenum Press, New York, pp. 241-269 (1979).

[42] I.J. Leitch, M.D. Bennett, Genome downsizing in polyploid plants, Biol. J. Linn. Soc. 82 (2004) 651–663.

[43] C. Parisod, K. Alix, J. Just, M. Petit, V. Sarilar, C. Mhiri, M. Ainouche, B. Chalhoub, M.A. Grandbastien, Impact of transposable elements on the organization and function of allopolyploid genomes, New Phytol. 186 (2010) 37–45.

[44] L.H. Rieseberg, C. Randal-Linder, G.J. Seiler, Chromosomal and genic barriers to introgression in *Helianthus*, Genetics 141 (1995) 1163-1171.

[45] B. McClintock, The significance of responses of the genome to challenge, Science 226 (1984) 792-801.

[46] J.J. Doyle, L.E. Flagel, A.H. Paterson, R.A. Rapp, D.E. Soltis, P.S. Soltis, J.F. Wendel, Evolutionary genetics of genome merger and doubling in plants, Annu. Rev. Genet. 42 (2008) 443–461.

[47] M. Freeling, M.R. Woodhouse, S. Subramaniam, G. Turco, D. Lisch, J.C. Schnable, Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants, Curr. Opin. Plant Biol. 15 (2012) 131–139.

[48] C. Parisod, N. Senerchia, Responses of transposable elements to polyploidy. In: Plant transposable elements (eds. M.A. Grandbastien, J.M. Casacuberta), Springer, Berlin, Germany, pp. 147–148 (2012).

[49] M. Petit, C. Guidat, J. Daniel, E. Denis, E. Montoriol, Q.T. Bui, K.Y. Lim, A. Kovarik, A.R. Leitch, M.A. Grandbastien, C. Mhiri, Mobilization of retrotransposons in synthetic allotetraploid tobacco, New Phytol. 186 (2010) 135–147.

[50] K. Kashkush, M. Feldman, A.A. Levy, Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat, Nature Genet. 33 (2003) 102–106.

[51] Y.M. Wang, Z.Y. Dong, Z.J. Zhang, X.Y. Lin, Y. Shen, D. Zhou, B. Liu, Extensive *de novo* genomic variation in rice induced by introgression from wild rice (*Zizania latifolia* Griseb.), Genetics 170 (2005) 1945–1956.

[52] J. Zou, D. Fu, H. Gong, W. Qian, W. Xia, C. Pires, R.Y. Li, Y. Long, A.S. Mason, T.J. Yang, Y.P. Lim, B.S. Park, J. Meng, *De novo* genetic variation associated with retrotransposon activation, genomic rearrangements and trait variation in a recombinant inbred line population of *Brassica napus* derived from interspecific hybridization with *Brassica rapa*, Plant J. 68 (2011) 212-224.

[53] R.C. Paz, A.P. Rendina Gonzalez, M.S. Ferrer, R.W. Masuelli, Short-term hybridisation activates *Tnt1* and *Tto1 Copia* retrotransposons in wild tuber-bearing *Solanum* species, Plant Biol. 17 (2015) 860-869.

[54] N. Senerchia, F. Felber, C. Parisod, Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation, Proc. R. Soc. B 282 (2015) 20142874.

[55] S. Renny-Byfield, M. Chester, A. Kovařík, S.C. Le Comber, M.A. Grandbastien, M. Deloger, R.A. Nichols, J. Macas, P. Novák, M.W. Chase, A.R. Leitch, Next Generation Sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs, Mol. Biol. Evol. 28 (2011) 2843–2854.

[56] S.Y. Dillen, V. Storme, N. Marron, C. Bastien, S. Neyrinck, M. Steenackers, R. Ceulemans, W. Boerjan, Genomic regions involved in productivity of two interspecific poplar families in Europe. 1. Stem height, circumference and volume, Tree Genet. Genomes 5 (2009) 147–164.

[57] R. Monclus, E. Dreyer, M. Villar, F.M. Delmotte, D. Delay, J.M. Petit, C. Barbaroux, D. Le Thiec, C. Bréchet, F. Brignolas, Impact of drought on productivity and water use efficiency in 29 genotypes of *Populus deltoides* × *Populus nigra*, New Phytol. 169 (2006) 765-777.

[58] L. Sterck, S. Rombauts, S. Jansson, F. Sterky, P. Rouzé, Y. Van de Peer, EST data suggest that poplar is an ancient polyploid, New Phytol. 167 (2005) 165–170.

[59] L. Zeng, N. Zhang, Q. Zhang, P.K. Endress, J. Huang, H. Ma, Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets, New Phytol. 214 (2017) 1338-1354.

[60] G.A. Tuskan, S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R.R. Bhalerao, R.P. Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G.L. Chen, D. Cooper, P.M. Coutinho, J. Couturier, S. Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroeve, A. Déjardin, C. dePamphilis, J. Detter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A.

Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjärvi, J. Karlsson, C. Kelleher, R. Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J.C. Leplé, P. Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D.R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryaboy, J. Schmutz, J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C.J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. Van de Peer, D. Rokhsar, The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray), Science 313 (2006) 1596–1604.

[61] G.T. Slavov, S.P. DiFazio, J. Martin, W. Schackwitz, W. Muchero, E. Rodgers-Melnick, M.F. Lipphardt, C.P. Pennacchio, U. Hellsten, L.A. Pennacchio, L.E. Gunter, P. Ranjan, K. Vining, K.R. Pomraning, L.J. Wilhelm, M. Pellegrini, T.C. Mockler, M. Freitag, A. Geraldes, Y.A. El-Kassaby, S.D. Mansfield, Q.C.B. Cronk, C.J. Douglas, S.H. Strauss, D. Rokhsar, G.A. Tuskan, Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*, New Phytol. 196 (2012) 713–725.

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

[63] Z. Xu, H. Wang, LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons, Nucl. Acids Res. 35 (2007) W265–268.

[64] E.L. Sonnhammer, R. Durbin, A dot-matriprogram with dynamic threshold control suited for genomic DNA and protein sequence analysis, Gene 167 (1995) GC1–GC10.

[65] Z. Zhang, S. Schwartz, L. Wagner, W. Miller, A greedy algorithm for aligning DNA sequences, J. Comput. Biol. 7 (2000) 203-214.

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

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

[68] L. Natali, R.M. Cossu, F. Mascagni, T. Giordani, A. Cavallini, 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 (2015) 107.

[69] M. Buti, M. Moretto, E. Barghini, F. Mascagni, L. Natali, M. Brilli, A. Lomsadze, P. Sonego, L. Giongo, M. Alonge, R. Velasco, C. Varotto, N. Surbanovski, M. Borodovsky, J.A. Ward, K. Engelen, A. Cavallini, A. Cestaro, D.J. Sargent, The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry), GigaScience 7 (2018) 1-14.

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

[71] J.J. Doyle, J.L. Doyle, Isolation of plant DNA from fresh tissue, Focus 12 (1989) 13–15.

[72] J. Logemann, J. Schell, L. Willmitzer, Improved method for the isolation of RNA from plant tissues, Anal. Biochem. 163 (1987) 16-20.

[73] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.

[74] R.M. Cossu, M. Buti, T. Giordani, L. Natali, A. Cavallini, A computational study of the dynamics of LTR retrotransposons in the Populus trichocarpa genome, Tree Genet. Genomes 8 (2012) 61–75.

[75] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.

[76] R.M. Cossu, T. Giordani, A. Cavallini, L. Natali, High-throughput analysis of transcriptome variation during water deficit in a poplar hybrid: a general overview, Tree Genet. Genomes 10 (2014) 53–66.

[77] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-seq, Nature Meth. 5 (2008) 621–628.

[78] K.A. Baggerley, L. Deng, J.S. Morris, C.M. Aldaz, Differential expression in SAGE: accounting for normal between-library variation, Bioinformatics 19 (2003) 1477–1483.

[79] C. Alkan, B.P. Coe, E.E. Eichler EE, Genome structural variation discovery and genotyping, Nat. Rev. Genet. 12 (2011) 363–376.

[80] K. Swaminathan, K. Varala, M.E. Hudson, Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey, BMC Genomics 8 (2007) 132–145.

[81] M.I. Tenaillon, M.B. Hufford, B.S. Gaut, J. Ross-Ibarra, Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*, Genome Biol. Evol. 3 (2011) 219–229.

[82] E. Barghini, L. Natali, R.M. Cossu, T. Giordani, M. Pindo, F. Cattonaro, S. Scalabrin, R. Velasco, M. Morgante, A. Cavallini, The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome, Genome Biol. Evol. 6 (2014) 776-791.

[83] P. Neumann, A. Koblizkova, A. Navratilova, J. Macas, Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement, Genetics 173 (2006) 1047-1056.

[84] B. Piegu, R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice, Genome Res. 16 (2006) 1262-1269.

[85] S. Pinosio, S. Giacomello, P. Faivre-Rampant, G. Taylor, V. Jorge, M.C. Le Paslier, G. Zaina, C. Bastien, F. Cattonaro, F. Marroni, M. Morgante, Characterization of the poplar pan-genome by genome-wide identification of structural variation, Mol. Biol. Evol. 33 (2016) 2706–2719.

[86] M.C. Ungerer, S.C. Strakosh, Y. Zhen, Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation, Curr. Biol. 16 (2006) R872–873.

[87] K.Y. Lim, D.E. Soltis, P.S. Soltis, J. Tate, R. Matyasek, H. Srubarova, A. Kovarik, J.C. Pires, Z. Xiong, A.R. Leitch, Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae), PLoS ONE 3 (2008) e3353.

[88] V. Sarilar, P.M. Palacios, A. Rousselet, C. Ridel, M. Falque, F. Eber, A.M. Chèvre, J. Joets, P. Brabant, K. Alix, Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids, New Phytol. 198 (2013) 593– 604.

[89] H. Shaked, K. Kashkush, H. Ozkan, M. Feldman, A.A. Levy, Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13 (2001) 1749-1759.

[90] N. Chantret, J. Salse, F. Sabot, S. Rahman, A. Bellec, B. Laubin, I. Dubois, C. Dossat, P. Sourdille, P. Joudrier, M.F. Gautier, L. Cattolico, M. Beckert, S. Aubourg, J. Weissenbach, M. Caboche, M. Bernard, P. Leroy, B. Chalhoub, Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*), Plant Cell 17 (2005) 1033–1045.

[91] C. Parisod, A. Salmon, T. Zerjal, M. Tenaillon, M.A. Grandbastien, M. Ainouche, Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*, New Phytol. 184 (2009) 1003–1015.

[92] J.C. Pires, J.W. Zhao, M.E. Schranz, E.J. Leon, P.A. Quijada, Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (*Brassicaceae*), Biol. J. Linn. Soc. 82 (2004) 675–688.

[93] J.A. Udall, P.A. Quijada, T.C. Osborn, Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L., Genetics 169 (2005) 967–979.

[94] X. Zhao, Y. Si, R.E. Hanson, C.F. Crane, H.J. Price, D.M. Stelly, J.F. Wendel, A.H. Paterson, Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton, Genome Res. 8 (1998) 479–492.

[95] C. Mhiri, C. Parisod, J. Daniel, M. Petit, K.Y. Lim, F. Dorlhac de Borne, A. Kovarik, A.R. Leitch, M.A. Grandbastien, Parental transposable element loads influence their dynamics in young *Nicotiana* hybrids and allotetraploids, New Phytol. 221 (2019) 1619-1633.

[96] B. Liu, C.L. Brubaker, G. Mergeai, R.C. Cronn, J.F. Wendel, Polyploid formation in cotton is not accompanied by rapid genomic changes, Genome 44 (2001) 321-330.

[97] A. Baumel, M. Ainouche, R. Kalendar, A.H. Schulman, Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina* anglica CE Hubbard (*Poaceae*), Mol. Biol. Evol. 19 (2002) 1218-1227.

[98] M. Charles, H. Belcram, J. Just, C. Huneau, A. Viollet, A. Couloux, B. Segurens, M. Carter, V. Huteau, O. Coriton, R. Appels, Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat, Genetics 180 (2008) 1071-1086.

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

[100] Yamazaki M, Tsugawa H, Miyao A, Yano M, Wu J, Yamamoto S, Matsumoto T, Sasaki T, Hirochika H (2001) The rice retrotransposon Tos17 prefers low-copy-number sequences as integration targets. Mol Gen Genomics 265:336–344

Table 1

Hybrid to mean of the parents abundance ratio (ratios between average coverages of the hybrid and the mean of the parents) of different lineages of LTR-REs. Significant differences for each lineage are indicated by different letters ($p < 0.05$) according to Bonferroni's multiple comparison test.

Lineage	Nr. of elements	Hybrid to mean of the parents abundance ratio	Bonferroni's multiple comparison test
Unknown	43	2.218	a
Gypsy - Unknown	50	1.630	a, b, c, d, e
Chromovirus	174	1.298	b, c, d, e
TAR/Tork	90	1.261	b, c, d, e, f, h
Ale I/Retrofit	42	0.971	b, c, d, e, f, g, h, i
Ivana	104	0.864	d, e, f, g, h, i
AleII	122	0.772	e, f, g, h, i
Ogre	67	0.762	d, e, f, g, h, i
Athila	126	0.533	e, f, g, h, i

Table 2

Number of LTR-RE insertion sites shared or not among the parents (*P. deltoides* and *P. nigra*) and their hybrid PxC1. The ID code of the 11 full-length LTR-REs selected for this analysis were: PRT_Chr4_44 (1), PRT_Chr4_48 (2), PRT_Chr5_48 (3), PRT_Chr6_51 (4), PTR_Chr10_47 (5), PRT_Chr11_27 (6), PRT_Chr11_31 (7), PRT_Chr11_60 (8), PRT_Chr14_3 (9), PRT_Chr18_7 (10), PRT_Chr19_62 (11). X indicates the presence, 0 the absence of the site in each of the three genotypes. Hybrid to mean of the parents abundance ratio of each LTR-RE is also reported.

Insertion site present in:				Retrotransposon											
	Hybrid	*P. deltoides*	*P. nigra*	1	2	3	4	5	6	7	8	9	10	11	To-tal
1	X	X	X	46	25	1	2	1	1	1	8	5	0	1	91
2	X	X	0	23	22	1	2	0	0	0	3	0	1	0	52
3	X	0	X	53	31	1	5	2	1	0	4	8	1	0	106
4	0	X	X	8	7	0	0	0	0	0	3	0	0	1	19
5	0	X	0	49	42	0	3	0	1	1	12	1	1	1	111
6	0	0	X	70	52	1	6	1	1	0	9	8	0	0	148
7	X	0	0	5	1	0	2	0	0	0	0	0	0	1	9
Hybrid to mean of the parents abundance ratio				1.6	1.4	1.1	1.6	5.9	2.8	1.6	5.9	2.8	1.6	1.6	
Lineage				*TAR/ Tork*	*Iva-na*	*Ale II*	*TAR/ Tork*	*Chr-omo-virus*	*Chr-omo-virus*	*Ale II*	*TAR/ Tork*	*Chr-omo-virus*	*Ale II*	*Iva-na*	

LEGENDS FOR FIGURES

Fig. 1. Procedures used (a) to identify LTR-RE insertions by using high-coverage Illumina gDNA sequencing and to validate them by bioinformatics analysis (b) or by PCR (c).

Fig. 2. Composition of the dataset of 828 full-length LTR-REs of *P. trichocarpa* isolated in this study. The number of elements of each *Copia* and *Gypsy* lineage are reported on the left and on the right, respectively.

Fig. 3. Relationships between average coverages of LTR-REs of *P. deltoides* and *P. nigra* (left) and of the PxC1 hybrid and the mean of its parents (right).

Fig. 4. Electrophoretic patterns of PCR products obtained using primers designed to amplify the insertion site or an internal portion of three LTR-REs in *P. deltoides* (Pd), *P. nigra* (Pn) and their hybrid (H). Molecular weights of amplified bands are reported (in bp) on the right of each gel.

Fig. 5. Relationship between RPKMs of LTR-REs of *P. deltoides* and *P. nigra*.

Fig. 6. RPKM of 14 LTR-REs over- or under-expressed in the hybrid compared to the mean of the parents. Asterisks indicate the significance of the difference, after Bonferroni's correction (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

Supplementary Materials

Suppl. Materials 1.

List of primers used to validate new insertions of LTR-REs in the genome of *P. deltoides* x *P. nigra*.

File Excel

Suppl. Materials 2.

Alignments of Illumina reads of the three genotypes used in this study (the hybrid and its parents) to ten fragments of poplar single copy genes.

Suppl. Materials 3.

List of 828 full-length LTR-REs of *P. trichocarpa*, with annotation. The average coverage and the RPKM in the three genotypes analyszed (*P. deltoides*, *P. nigra*, and their hybrid) are also reported.

File Excel

Suppl. Materials 4.

The 27 nt-long fragments at the 5'end of the 5'-LTR of 11 analyszed elements (including the dinucleotide TG).

>PTR_Chr10_47

TGTCACAGCCTTAGGCACGAGGGCGTG

>PRT_Chr11_27

TGATGCAGGCTAGAACGGAACTCACAT

>PRT_Chr19_62

TGTTAAAGTTGTTAGGATTTCTTTGAA

>PRT_Chr11_31

TGTCTAATTGGCTGAATGAATACCTCT

>PRT_Chr6_51

TGTCACAGTGTCAAAAGTGCACGAGGT

>PRT_Chr4_44

TGTTGTGCATACTGGACCGAAAGCAAA

>PRT_Chr4_48

TGAGAAATAATTAGGAGGCTTAACCTA

>PRT_Chr14_3

TGTTGCAATTGTCAACATTTTTGTCAA

>PRT_Chr5_48

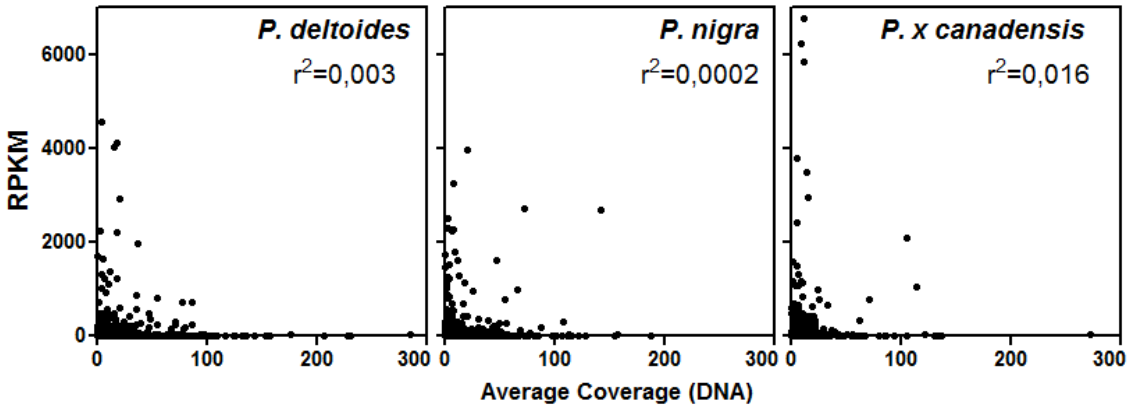
TGGAATCTACAACACTCTACATCTATA

>PRT_Chr11_60

TGTAGCAAATTGTCAACATTTGCAATT

>PRT_Chr18_7

TGTAGCAAATTGTCAACATTTTCTCAA



Supplementary Figure 1. Relationship between RPKM expression values of each element of the 828 *P. trichocarpa* LTR-REs and the respective number of mapped DNA reads in *P. deltoides*, *P. nigra* and their hybrid.

Suppl. Table 1. RPKM and average coverage of the 7 LTR-REs over-expressed in the hybrid than as obtained calculating the means of parents

LTR-RE ID	Super-family	Lineage	RPKM		Average coverage	
			Hybrid	Mean of the parents	Hybrid	Mean of the parents
PRT_Chr04_15	*Gypsy*	*Athila*	18.51	3.98	13.81	28.69
PRT_Chr13_26	*Gypsy*	*Athila*	115.50	35.68	10.57	19.67
PRT_Chr03_3	*Gypsy*	*Ogre*	69.37	19.83	13.90	14.73
PRT_Chr12_31	*Gypsy*	*Ogre*	77.24	12.36	17.93	17.53
PRT_Chr16_3	*Gypsy*	*Ogre*	427.52	146.60	12.54	13.41
PRT_Chr04_42	*Gypsy*	*Chromovirus*	21.75	2.65	46.98	54.54
PRT_Chr01_69	*Copia*	*Ivana*	81.39	10.07	5.15	12.19

Suppl. Table 2. Mean hybrid to mean of the parents RPKM ratio of different lineages of LTR-REs. Significant differences for each lineage are indicated by different letters ($p < 0.05$) according to Bonferroni's multiple comparison test.

Lineage	Nr. of elements	Hybrid to mean of the parents RPKM ratio	Bonferroni's multiple comparison test
Chromovirus	174	1.681	a
Unknown	43	1.633	a
Athila	126	1.620	a
Ale I/Retrofit	42	1.586	a
Gypsy – Unknown	50	1.500	a
AleII	122	1.307	a
Ivana	104	1.282	a
Ogre	67	1.141	a
TAR/Tork	90	1.076	a