



HAL
open science

Specificities and Dynamics of Transposable Elements in Land Plants

Corinne Mhiri, Filipe Borges, Marie-Angèle Grandbastien

► **To cite this version:**

Corinne Mhiri, Filipe Borges, Marie-Angèle Grandbastien. Specificities and Dynamics of Transposable Elements in Land Plants. *Biology*, 2022, 11 (4), pp.488. 10.3390/biology11040488 . hal-04179211

HAL Id: hal-04179211

<https://hal.inrae.fr/hal-04179211>

Submitted on 22 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Review

Specificities and Dynamics of Transposable Elements in Land Plants

Corinne Mhiri * , Filipe Borges  and Marie-Angèle Grandbastien 

Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000 Versailles, France; filipe.borges@inrae.fr (F.B.); marie-angele.grandbastien@inrae.fr (M.-A.G.)

* Correspondence: corinne.mhiri@inrae.fr

Simple Summary: Transposable elements are dynamic components of plant genomes, and display a high diversity of lineages and distribution as the result of evolutionary driving forces and overlapping mechanisms of genetic and epigenetic regulation. They are now regarded as main contributors for genome evolution and function, and important regulators of endogenous gene expression. In this review, we survey recent progress and current challenges in the identification and classification of transposon lineages in complex plant genomes, highlighting the molecular specificities that may explain the expansion and diversification of mobile genetic elements in land plants.

Abstract: Transposable elements (TEs) are important components of most plant genomes. These mobile repetitive sequences are highly diverse in terms of abundance, structure, transposition mechanisms, activity and insertion specificities across plant species. This review will survey the different mechanisms that may explain the variability of TE patterns in land plants, highlighting the tight connection between TE dynamics and host genome specificities, and their co-evolution to face and adapt to a changing environment. We present the current TE classification in land plants, and describe the different levels of genetic and epigenetic controls originating from the plant, the TE itself, or external environmental factors. Such overlapping mechanisms of TE regulation might be responsible for the high diversity and dynamics of plant TEs observed in nature.

Keywords: transposable element; transposition control; plant genome; TE classification; evolution; stress; silencing; epigenetic; methylation; polyploidy



Citation: Mhiri, C.; Borges, F.;

Grandbastien, M.-A. Specificities and Dynamics of Transposable Elements in Land Plants. *Biology* **2022**, *11*, 488. <https://doi.org/10.3390/biology11040488>

Academic Editor: Silke Jensen

Received: 11 February 2022

Accepted: 18 March 2022

Published: 23 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Transposable elements (TEs) can be defined as repetitive DNA sequences able to move/transpose throughout their host genome. They were first discovered in maize by Barbara McClintock in the 1940s as controlling elements able to modify gene expression and change their location upon genomic stress, such as chromosomal double-strand breaks [1]. With the development of molecular biology and sequencing technologies, the detection of mobile genetic elements has been generalized to almost all living organisms. Their high abundance in some genomes (>80% in maize) [2] and extreme diversity in transposition modes and insertion profiles resulted in a progressive interest of the research community for studying the biology of these repetitive sequences, and the way they interact and coevolve with their host genome. Initially seen as invading parasitic sequences because of their proliferative and mutational abilities [3,4], TEs have been progressively considered as important components of eukaryotic genomes since the discovery of some 'useful' TEs contributing to gene expression regulation or enzymatic functions [5,6]. Such opinion changes, considering TEs as 'facilitators of evolution' for their host organisms, have been well documented [7–10].

After presenting a general updated picture of TE composition and diversity in plants, this review will also focus on the different control levels acting in plants to regulate TE expression, activity and expansion, thus contributing to plant evolution and adaptation.

2. Plant TE Landscape

2.1. A Highly Variable TE Abundance

Due to their ubiquity, abundance and transposition activity, TEs have been proposed as major contributors to genome size, along with other mechanisms such as recombination rate and polyploidization [11]. Among living organisms, land plants (regrouping Bryophytes, Pteridophytes, and seed plants), and especially flowering plants (Angiosperms), display one of the largest genome size variability exceeding 2400-fold, with C-values ranging from 0.07 pg (65 Mb/1 C) for the small carnivorous plant *Genlisea tuberosa* to 152.23 pg (149 Gb/1 C) for the monocot lily species *Paris japonica* (<https://cvalues.science.kew.org/> accessed on 10 March 2022). Indeed, TEs seem to account for the variable proportion of plant genomes sequenced to date, spanning from ~3% in the small 82 Mb carnivorous *Utricularia gibba* [12] to ~85% in allohexaploid wheat (*Triticum aestivum*) [13] or maize genome [2].

2.2. Challenging Evaluation of Plant TE Diversity and Classification

Our knowledge about TEs structure, organization and transposition mechanisms has greatly progressed since the discovery of the first TE sequences in plants (the maize Ac/Ds elements) in 1984 by two different laboratories [14,15]. Transposable elements are extremely diverse and use various mechanisms to move. The so-called autonomous elements encode all specific functions to achieve their mobility, while some non-autonomous elements hitch-hike mobility proteins from autonomous copies, or from other TEs to transpose.

Several types of TEs might co-exist in a given genome, and each TE type can harbor multiple TE copies clustered into different families according to their sequence similarity. As some transposition mechanisms are prone to generate mutations, each family might have evolved over time, displaying a continuum of more or less diverged copies composed of both autonomous and defective elements [7,8,10,16]. Such behaviour made TE identification and classification a difficult task.

In 1989, the first TE nomenclature organized TEs into two classes according to their transposition intermediate [17]: (1) an RNA intermediate for retrotransposons (class I elements) that move via a replicative “copy-and-paste” mechanism, where a “mother” copy gives rise to several “daughter” copies without excising itself; (2) a DNA intermediate for DNA transposons (class II elements) that use a conservative “cut-and-paste” mechanism for their transposition, where the “mother” copy excises from its location to insert elsewhere in the genome (“jumping gene” concept).

This bimodal schematic classification has been further refined in 2007 (1) by creating subclasses in the class II group to include TEs that use a DNA replicative mechanism, such as Helitrons, and (2) by setting the “80-80-80” rule to define the identity percentage TE copies should share to belong to the same family, i.e., sequences over 80 bp sharing at least 80% sequence identity in at least 80% of their internal domain or terminal repeats (or both) [16,18,19]. This TE hierarchical classification splits TEs into the two previously cited classes (Class I and Class II), then into subclasses, orders and superfamilies [18]. This classification is based on the presence and order of coding regions for specific proteins or structural motifs present in TE sequences, and on their transposition specificities (presence and sequence of target site duplications (TSDs), i.e., short 2–11 bp DNA duplications, generated upon TE transposition at the new insertion site). For example, the presence of long terminal repeats (LTRs) in direct orientation at TE 3' and 5' ends is the signature of LTR-retrotransposons (LTR-RTs). The occurrence of the reverse transcriptase (RT) domain in TE ORF is a hallmark of most but not all Class I retrotransposons. Like self-replicating entities as viruses, TEs have a modular evolution, exchanging essential or facultative protein-coding domains that may blur TE classification [19].

If these first four classification levels are generally well accepted, the need of extra levels to reach the TE family level is still unclear and may vary according to the TE considered. For example, the superfamily subgroups “chromovirus” or “non-chromovirus” have been introduced for Gypsy LTR-RTs [20]. An increase in family/lineage number or

even new superfamilies will probably arise with the accumulation of genome sequencing data and the improvement of bioinformatic pipelines for TE detection and annotation. This hierarchical classification also does not include some non-autonomous TEs, i.e., TEs that are still able to transpose but need extra-function in trans coming from another TE element (phylogenetically related or not). This includes LARDs (large retrotransposon derivatives), TRIMs (transposon in miniature) and SMARTs (small LTR-retrotransposon) for Class I TEs, and MITEs (miniature inverted-repeats transposable elements), SNACs (small non-autonomous CACTA) or MULEs (Mu-like elements) for class II TEs. Many such non-autonomous elements share enough sequence similarity/motifs to be easily linked to a candidate “helper” autonomous TE family, as reported in rice with the isolation of the complete RIRE2 Gypsy retrotransposon displaying a high LTR sequence similarity to sequence extremities of the defective *Dasheng* retrotransposon [21]. Another example is the description of both complete and truncated—but nevertheless active—CACTA *Caspar* elements in Triticeae [16,22]. Some other non-autonomous TEs may not present any clear feature allowing them to be classified, as some non-autonomous short TIR-harboring TEs that may share only a few bases homology with the autonomous helper [16]. In the classification we propose, we choose to include the non-autonomous SINES as a full order, as these TEs do not correspond to deleted versions of autonomous class I TEs.

Table 1 and Figure 1 present, respectively, the up-to-date classification and structures of plant transposable elements adapted from Wicker et al. [16]. It is important to note that only a subset of existing TE superfamilies in all living organisms (as reported in rebase <https://www.girinst.org/rebase/update/browse.php> accessed on 10 March 2022) has been detected in land plants (~30% of class II superfamilies, and ~17% of class I—representing 20% of all described superfamilies). Plant TEs fall into six different orders, four orders corresponding to class I elements (LTR-RTs, *Penelope*-like elements (PLE), long interspaced nuclear elements (LINEs), short interspaced nuclear elements (SINEs)) and two orders including elements of class II (terminal inverted repeat (TIR) transposons, Helitrons) (Table 1).

Table 1. Plant transposable element (TE) classification compiled from [16] with updates from [23] for Copia lineages, ref. [20] for Gypsy LTR retrotransposons (LTR-RTs), ref. [24–26] for *Penelope*-like elements (PLEs), ref. [27] for long interspaced nuclear elements (LINEs), ref. [28,29] for short interspaced nuclear elements (SINEs), and [30,31] for Sola elements.

Class	Order (Non-Autonomous TE Name)	Superfamily	Family/Lineage	Plant Family Examples
Class I (retrotransposons)	LTR-Retrotransposons (LARD) (TRIM/SMART)	Copia	Osser	<i>Volvox canteri</i> Osser
			Bryco	representatives in moss species
			Lyco	representatives in clubmosses species (<i>Lycopodiaceae</i>)
			Gymco-I	representatives in gymnosperms species
			Gymco-II	representatives in gymnosperms species
			Gymco-III	representatives in gymnosperms species
			Gymco-IV	representatives in gymnosperms species
			Ale/Retrofit	<i>Oryza longistaminata</i> Retrofit, <i>Oryza sativa</i> Hopscotch <i>Oryza sativa</i> Oryko1-1 and Ilona, <i>Hordeum vulgare</i> HORPIA, <i>Nicotiana tabacum</i> Queenti

Table 1. Cont.

Class	Order (Non-Autonomous TE Name)	Superfamily	Family/Lineage	Plant Family Examples
			Ikeros	<i>Zea mays</i> Sto-4
			Tork	<i>Nicotiana tabacum</i> Tnt1, Tto1 and Tnt2, <i>Solanum lycopersicum</i> Tork4, <i>Ipomea batatas</i> Batata
			Alesia	low copy number representatives in many Angiosperms, close to the Ale lineage
			Angela	<i>Triticum aestivum</i> Angela, <i>Oryza sativa</i> RIRE1, <i>Hordeum vulgare</i> BARE1
			Bianca	<i>Triticeae</i> Bianca, <i>Arabidopsis thaliana</i> RomaniAT5
			SIRE/Maximus	<i>Solanum lycopersicum</i> ToRTL1, <i>Zea mays</i> Opie-2, <i>Glycine max</i> SIRE1
			TAR	<i>Oryza</i> spp. Houba and Osr-1, <i>Arabidopsis thaliana</i> ATcopia95
		Gypsy (Chromovirus)	Galadriel	<i>Solanum esculentum</i> Galadriel, <i>Musa</i> Monkey, Tntom1
			Tekay	<i>Hordeum vulgare</i> Bagy-1, <i>Arabidopsis thaliana</i> Legolas Peabody, <i>Oryza sativa</i> RIRE3, <i>Lilium henryi</i> Del
			Reina	<i>Zea mays</i> Reina, <i>Arabidopsis thaliana</i> Gloin or Gimli
			CRM	<i>Zea mays</i> CRM (centromeric retrotransposon of maize), <i>Beta</i> <i>vulgaris</i> Beetle1, <i>Oryza sativa</i> RIRE7
		(Non-chromovirus)	Phygy	<i>Phycomitrella patens</i> Chr21 (4035670,4045566)
			Selgy	<i>Selaginella moellendorffii</i> LTR-RT
			Athila	<i>Arabidopsis thaliana</i> Athila4-1, Diaspora, <i>Hordeum vulgare</i> Bagy-2
			TatI	<i>Selaginella moellendorffii</i> LTR-RT
			TatII	<i>Picea abies</i> , <i>Picea glauca</i> LTR-RTs
			TatIII	<i>Picea abies</i> , <i>Picea glauca</i> LTR-RTs
			Ogre/TatIV + TatV	<i>Pisum sativum</i> Ogre <i>Zea mays</i> Cinfu1-1, <i>Arabidopsis</i> <i>thaliana</i> Tat4-1, <i>Oryza sativa</i> RIRE2,
			Retand/TatVI	<i>Sorghum bicolor</i> RetroSor1, <i>Silene latifolia</i> Retand <i>Pinus taeda</i> (loblolly pine) and <i>Picea</i> <i>abies</i> (Norway spruce) <i>Dryad</i> PLEs by horizontal transfer
	Non-LTR retrotransposons PLE	<i>Penelope/Poseidon</i>		<i>Selaginella moellendorffii</i> spike moss, <i>Pinus taeda</i> and <i>Picea abies</i> EN(-)PLEs sweet potato Llb, <i>Beta vulgaris</i> BvL1
	LINE	EN(-)PLE L1	Llb	<i>Cannabis sativa</i> LINE-CS, <i>Beta</i> <i>vulgaris</i> Belline2, Belline5
			LINE-CS	<i>Beta vulgaris</i> Belline1/BNR
			BNR	<i>Carica papaya</i> L1-26_Cpa, <i>Solanum</i> <i>tuberosum</i> L1-3_Stu, <i>Vitis vinifera</i>
			PUR	<i>Zea mays</i> Cin4
			Cin4	<i>Oryza sativa</i> Karma
			Karma	<i>Oryza sativa</i> LINE-1 or OSLINE1-4, <i>Zea mays</i> L1-2_ZM
			nubo	<i>Malus x domestica</i> RTE-1_Mad, <i>Solanum tuberosum</i> RTE-1_Stu
		RTE	plant RTE	

Table 1. Cont.

Class	Order (Non-Autonomous TE Name)	Superfamily	Family/Lineage	Plant Family Examples
Class II Subclass 1	SINE	tRNA		<i>Nicotiana tabacum</i> TS, Au, Solanales SoLS-II, Brassicaceae BraS-I, SB families, mainly found in Angiosperm
	TIR (MITE)	Tc1-Mariner		Stowaway (MITE): <i>Sorghum bicolor</i> Stowaway, Brassicaceae BraSto
Subclass 2	(MULE)	hAT		<i>Zea mays</i> Ac/Ds, <i>Antirrhinum majus</i> Tam3, <i>Nicotiana tabacum</i> Slide
		Sola		<i>Physcomitrella</i> Sola1, found also in <i>Capsicum annuum</i> and <i>C. baccatum</i>
		MuDR-Foldback		<i>Zea mays</i> Mu, MULEs
		PIF-Harbinger		<i>Zea mays</i> PIFa, <i>Oryza sativa</i> Pong; Tourist (MITE): mPing/Ping; mPIF/PIFa
		CACTA		<i>Zea mays</i> En/Spm, <i>Arabidopsis thaliana</i> CAC1, <i>Antirrhinum majus</i> Tam1, <i>Petunia hybrida</i> PsI
	Helitron	Helitron		<i>Oryza sativa</i> , <i>Arabidopsis thaliana</i> AthE1 Atrep, <i>Ipomoea tricolor</i> Hel-It1

In plants, class II elements (subclass 1) belonging to the TIR order (also called DNA transposons) fall into six superfamilies, based on the structure and sequence of their transposase and on the sequence of their terminal inverted repeats (TIRs) (Figure 1). Transposase is the protein catalyzing their transposition, while TIRs harbor key sequences recognized by the DNA-binding domains of the transposase during a transposition event. Some TIR elements also harbor additional coding sequences, as the maize MuDR, and plant CACTA or PIF/Harbinger elements [16]. Most of these superfamilies are also characterized by specific target site duplications (TSDs) lengths, generated after the filling of DNA nicks generated by the transposase on the integration site. Their transposition is not always strictly conservative and could lead to an increase of copy number if it occurs before a DNA replication fork [32].

Replicating plant TEs fall into two major groups: (1) Helitrons (class II-subclass 2, see Table 1) replicate through a rolling-circle (RC) mechanism from one DNA strand, without generating TSDs, by using a RepHel protein with a RC replication initiator (Rep) and DNA helicase (Hel) domains, in association with an ssDNA-binding “replication protein A” (RPA) [33]. (2) Class I retrotransposons replicate from RNA templates by reverse transcription using a TE-encoded reverse transcriptase (RT) and use at least one additional protein to mediate their insertion into their host genome, such as endonuclease (EN) or DDE integrase (INT). We do not include the DIRS superfamily as a member of land plants, as DIRS elements have only been found in green algae until now [34].

Among the four retrotransposon orders present in plants, SINEs occupy a particular place, as these small non-coding and non-autonomous elements of a few hundred base pairs exploit the transposition machinery of LINES to ensure their amplification. Plant SINEs are derived from tRNAs [28]. They are transcribed by polymerase III, harbor short degenerated internal promoters (A and B boxes), and display mostly A tail at the 3'-end. Apart from these small structural domains, SINEs display a high sequence diversity that hinders their detection and characterization. Recently, a 37 pb Angio-domain located in the 3'-end has been reported in many Angiosperm SINEs [29].

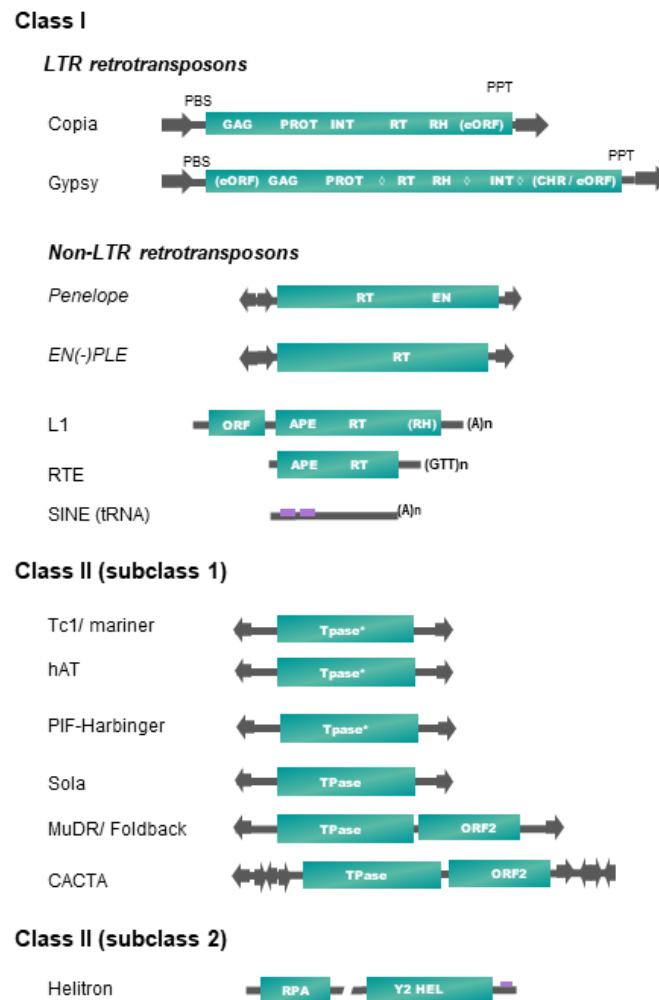


Figure 1. Structure and organization of plant transposable element superfamilies (adapted from [16]). Schemes are not to scale. Protein coding domains: APE = apurinic endonuclease, CHR = chromodomain, EN = endonuclease, GAG = capsid protein, HEL = helicase, INT = integrase, PROT = proteinase, RH = RNase H, RPA = replication protein A, RT = reverse transcriptase. eORF = extra open reading frame (unknown function), TPase = transposase (* with DDE motif), YR = tyrosine recombinase, Y2 = YR with YY motif, \diamond = different possible locations of an additional cellular-like ribonuclease H (aRH) specific of the Tat lineages (see Table 1). Optional protein-coding domains only present in some superfamily lineages are indicated in brackets. Some structural features are also represented. Terminal repeats in the same or reverse orientation are indicated by black arrows, and purple rectangles refer to diagnostic sequences present in non-coding sequences. Specific base termination of some TEs are also indicated. PBS = primer binding site, PPT = poly purine tract. Interrupted line in Helitron representation means that the region may contain one or more additional ORFs.

The second non-LTR retrotransposon order present in plants, LINES, contains elements belonging to the L1 and the RTE (retrotransposable element) superfamilies (Table 1 and Figure 1), which are two of the five known superfamilies of LINES detected in eukaryotes [16]. RTE and L1 LINES have one or two open reading frames (ORFs), respectively, and code for proteins required for retrotransposition, such as an endonuclease (EN), a RT, and often a ribonuclease H (RNase H (RH)). The L1 ORF1 is involved in the binding, protection and transport of the RNA intermediate used for retrotransposition. At their 3'-end lies a stretch of (A)_n for L1 or (GTT)_n for RTE involved in the reverse transcription initiation. A recent study shows that plant LINES extracted from 23 genomes fall into only seven L1 and one RTE families/lineages/subclades [27]. As the reverse transcription starts from the

3'-end of LINES and does not always reach the 5'-end, many incomplete daughter copies can be generated.

Between their bordering direct repeats 5'-LTR and 3'-LTR, autonomous LTR-retrotransposons (LTR-RTs) code for structural capsid-like (GAG) and functional (POL) proteins needed for their retrotransposition cycle (RT = reverse transcriptase, RH, INT = integrase), resembling the replication cycle of retroviruses. Only two out of the five superfamilies found in eukaryotes are represented in plants [16]. Plant LTR-RTs are further classified into Copia/Ty1 or Gypsy/Ty3 superfamilies according to the order of their coding *pol* domains. Recently, a systematic survey of plant LTR-RTs in 80 plant genomes refined the classification by introducing 16 lineages/families into the Copia/Ty1 superfamily and 14 lineages/families into the Gypsy/Ty3 group (six with a chromo-domain and eight without) [20]. Two Gypsy lineages, Chlamyvir and Tcn1, having only representatives in algae and non-*Viridiplantae* species, have not been included in Table 1. Non-autonomous derivatives of variable size (from a few hundred bp up to 25 kb) have been characterized in plants, containing between both LTRs a DNA sequence of variable length, either non-coding or reminiscent of some retrotransposon internal domains. Large internal sequences (>4 kb) define LARDs (large retrotransposon derivatives), and short ones (<4 kb) are often called TRIM (terminal repeat retrotransposon in miniature) [16].

Retrotransposons belonging to the *Penelope*-like elements (PLE) order are also found in some plant genomes, but with a patchy distribution. PLE encode an RT domain related to telomerase, a highly specialized class of non-mobile RTs responsible for chromosome end maintenance in most eukaryotes. Some PLEs also carry a second EN domain with a specific GIY-YIG motif. PLEs are bordered by repeats in direct or reverse orientation and are often subjected to 5' truncation upon retrotransposition, as non-LTR retrotransposons. EN(+)-PLEs (*Dryads* elements belonging to the *Penelope/Poseidon* group) and EN(-)-PLEs have been found in some Conifer genomes (Table 1, Figure 1), and some of them were presumably derived from a horizontal transfer (HT) event [26].

The accumulation of sequenced genomes and TE detection pipelines allow the analysis and comparison of TE composition and diversity across plant genera. Figure 2 presents a heatmap of genomic percentage of four types of TEs—LTR-RTs, LINES, SINES and TIR DNA transposons—across 74 Angiosperm species displaying variable genome sizes (Data collected from [35] in the Supplementary Tables S1 and S2 of this article). Among the different types of TEs, LTR-RTs (Copia/Ty1 and Gypsy/Ty3) occupy the largest proportion of these genomes, the highest being up to 80% in *Zea mays*. Such an increase can result from rapid amplification of only a few families. For example, *Oryza australiensis* has undergone a recent burst of transposition involving only three families (one Copia = RIRE1 and two Gypsy = Wallabi and Kangourou), which compose 60% of its genome [36]. Genomes of the legume tribe *Fabeae* are also dominated by the Ty3/Gypsy Ogre family/lineage, that accounts for 57% of genome size variation on average in this clade [37]. The predominance of one type of LTR-RT may vary depending on the taxa considered. For example, in *Gossypium* species, Copia LTR-RTs have accumulated in the small genome of *G. raimondii* (880 Mb), while Gypsy LTR-RTs (mainly Gorge3) have proliferated in large genomes lineages of *G. herbaceum* (1667 Mb) and *G. exiguum* (2460 Mb) [38]. Plant species belonging to the same order (see Brassicales, Poales, Figure 2) might display different TE compositions and genome sizes. Some plant species also harbor specific TE composition as shown in Figure 2, with dominance of LINES retrotransposons or TIR DNA transposon for the aquatic plant coontail *Ceratophyllum demersum* (33.6% of total genome size) and the small herbaceous plant *Trichopus zeylanicus* (~27.3% of total genome size). Non-LTR-retrotransposons have been shown to be abundant (~11.7%) in other plant genomes, such as *Arachis ipaensis*, one of the peanut parental genomes [39].

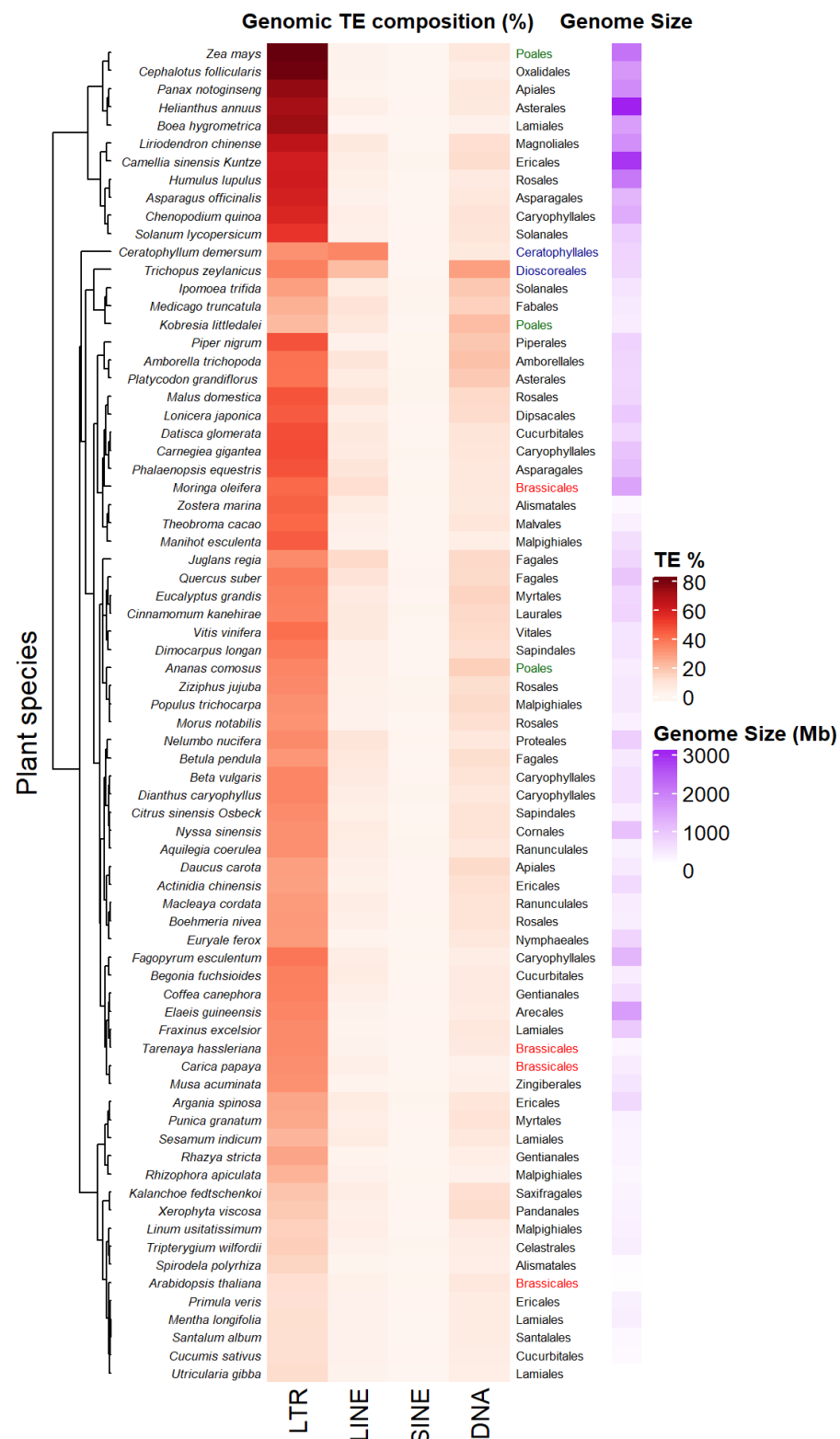


Figure 2. Transposable elements (TE) profiles in some land plant genomes. Species are clustered according to their TE profiles. TE percentages, plant orders and genome size estimations from 74 land plant species have been collected from [35] (data collected in Supporting information, Tables S1 and S2 from [35]). Some plant orders as Poales or Brassicales have been highlighted in colors (green and red respectively) in order to underline the diversity of TE composition between species belonging to the same plant order. Plant belonging to different orders as Dioscoreales and Ceratophyllales (in blue) can share close TE composition.

What are the forces taking place in the generation of such variable TE patterns in plants? The following section will present some of the factors known to influence TE diversity and dynamics in plants at the molecular level.

3. Regulating Factors of TE Transposition Control at the Molecular Level

The origin of species-specific differences in TE composition is still poorly understood. The TE profile observed in a given plant species may be the result of multiple regulating sources coming from TE themselves (regulatory motifs, biology) and host plant characteristics (epigenetic control, genome size, polyploidy level, sequence elimination mechanisms and reproductive system). All these evolutionary forces combined may lead to an equilibrium status, as illustrated in Figure 3 and discussed in the following paragraphs.

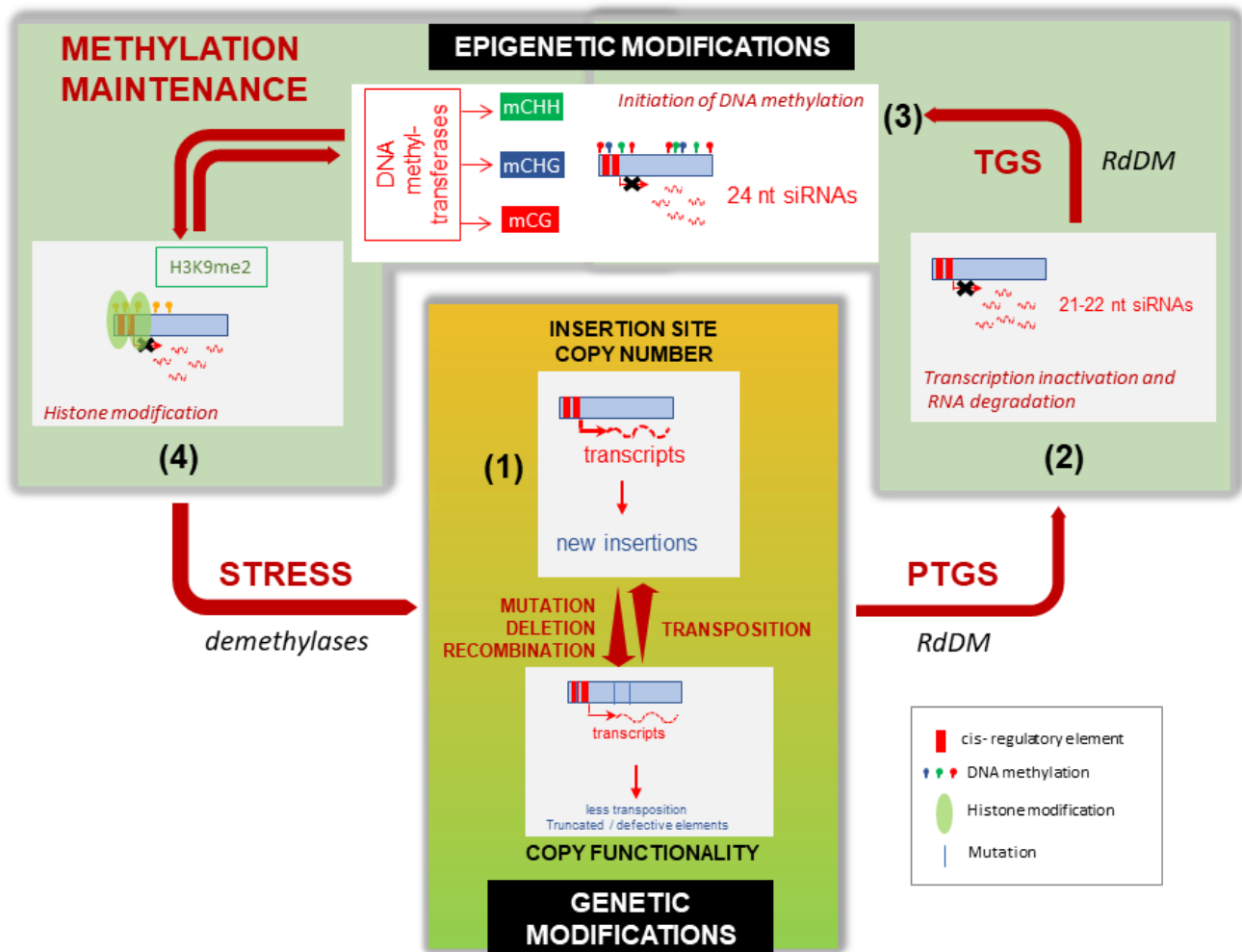


Figure 3. Different levels of TE transposition control in plants. (1) The activity of a new transposable element (TE) landing in a genomic place will depend on the presence of activating cis-regulating elements sensitive to stress, on the presence of similar copies, as well as insertion site preference (see text). Over time, TE copies are subjected to genetic modifications (mutations, deletions, recombinations) that will decrease copy functionality and number, as well as transposition ability. (2) High transcript abundance and presence of aberrant mRNA of inserted copies of a given TE trigger host post-transcriptional gene silencing (PTGS) that will selectively inactivate and degrade TE transcripts via the production of small interfering 21–22-nt RNAs. (3) Such inactivation is relayed by transcriptional gene silencing (TGS), relying on the production of 24-nt siRNA that will induce cytosine methylation in the three contexts, CHG, CHH and CG, via different DNA methyl-transferases (see text). (4) DNA methylation is actively

maintained over cell generations and time by histone methylation (mainly histone H3 lysine 9 dimethylation) and chromatin modification by a retro- feedback mechanism involving PolV non-coding transcripts (see text). Such epigenetic modifications can be removed by cell stress, modifying the epigenetic marks and allowing transient TE activity.

3.1. TE Regulatory Motifs Involved in TE Transcription/Activation

Despite the need for host cell machinery, autonomous TEs express specific proteins to ensure their self-propagation. These elements harbor regulatory motifs to promote their transcription and regulation, such as promoters for RNA polymerase II (Pol II, or Pol III for SINEs), and polyadenylation signals. The location of such cis-regulating sequences depends on the TE considered, and may disappear in truncated non-autonomous TEs [40]: in LTR-RTs, such motifs are present in duplicates in both 5'- and 3'-LTRs (and are also conserved in solo-LTRs, i.e., LTR-RT deleted versions generated upon ectopic recombination between LTRs); in class II TIR DNA transposons, such sequences are located in the 5' part of the coding sequence between TIRs; in LINEs, they are located in the 5'-UTR (untranslated region) and in the 3'-UTR, and only in the 5'-UTR in SINEs (Pol III promoter). Transcription regulation is a key point, especially for class I TEs that use transcripts not only for protein synthesis, but also as a matrix for their reverse transcription. The analysis of transcriptional activity of 56 families of transposable elements in different maize organs and cell culture reveals a higher expression of Gypsy LTR-RTs compared to other classes. In maize, expression profiles in organs vary according to the type of transposable elements, and seems to originate from a few active copies [41].

Since the first report in 1985 of transposition of the maize Bs1 element following barley stripe mosaic virus infection [42], the activation of transposable elements under various environmental challenges has been extensively reviewed [43]. In sessile organisms like plants, the transcriptional activation of TEs has been reported to be induced by a wide array of activating stimuli including biotic and abiotic stresses, such as pathogens and their elicitors, tissue culture, wounding, temperature or drought, UV or X-ray irradiation [43–45]. Different regulatory motifs behaving as transcriptional activators and involved in TE transcriptional activation have been isolated in the LTR of active tobacco Tnt1 or Tto1 LTR-RTs [46–49]. Such cis-regulating elements present in LTRs may vary within a single retrotransposon family and may favor the transcription (first step of transposition) of some specific subfamilies depending on the type of stress applied, as exemplified for tobacco Tnt1 LTR-RTs subfamilies [50]. Such fine-tuned transcriptional regulation upon stress may favor the transposition and thus the impact of some TEs on plant evolution over time. Comparative studies of Tnt1-related elements in wild tomato species revealed a similar fast evolution of LTR regulatory sequences, as well as for the *Citrus sinensis* Tc1 and Tc2 LTR-RTs [43].

3.2. Epigenetic Control of TEs

In order to prevent the potentially deleterious consequences of excessive transposition, host organisms have developed several mechanisms to repress (silence) expression of their endogenous TEs. Such processes of heritable inactivation of gene expression without altering the DNA sequence belong to the domain of epigenetics.

Phenotypes associated with *Ac* or *Spm* transposon silencing have been extensively described in maize in the early 1960s by Barbara McClintock and other researchers (reviewed in [51]). TEs can be found in multiple epigenetic states, ranging from a transcriptionally active state leading to transposition, to a silenced state without any transcriptional activity. Such silencing can be family-specific [52] and heritable throughout generations, and may also depend on the genomic context of the insertion site (near/within gene, pericentromeric or centromeric location) [53].

For example, de novo silencing of a newly inserted autonomous TE (as represented in Figure 3, step 1) initiates generally at the post-transcriptional level (post-transcriptional gene silencing (PTGS)) and results in degradation of TE mRNAs or interference with their

translation (Figure 3-step2). Then, a second mechanism ensures that TE silencing is maintained at the transcriptional level (transcriptional gene silencing (TGS)), by blocking TE transcription via DNA methylation and chromatin modifications at the corresponding TE sequence (Figure 3-step3). Such pathways involve different players including 21- to 24-nucleotide (-nt) small interfering RNAs (siRNAs), cytosine methylation (in CG, CHG or CHH contexts, where H = A, T or C), histone methylation (mainly di-methylation on lysine 9 of histone H3-H3K9me2) and recruitment of numerous host-encoded proteins. These epigenetic silencing events have been thoroughly studied in the model plant *Arabidopsis thaliana* (for a recent review, see Zhang et al. [54]).

One of the key events to initiate epigenetic silencing is the production of double-stranded RNAs from abundant, coding or non-coding aberrant RNAs, that occurs following hairpin formation, partial hybridization of two complementary mRNA, or produced by RNA-DEPENDENT RNA POLYMERASES (as RDR2/RDR6 in *Arabidopsis*) [55]. Depending on their type and origin, double-stranded RNAs are cleaved by DICER-LIKE PROTEINS into siRNAs of different sizes (21/22-nt by DCL2/4, 24-nt by DCL3 in *Arabidopsis*), and loaded onto ARGONAUTE proteins (AGO1 to AGO10 in *Arabidopsis*) to target complementary RNAs [55]. Several silencing outcomes may occur depending on the silencing complex interacting with the paired AGO/siRNA considered: for example, the RNA-induced silencing complex (RISC) associated with the paired AGO1/21–22-nt siRNA or microRNAs (miRNAs) will target and cleave complementary mRNAs, triggering PTGS. In contrast, TGS requires the plant-specific Pol IV to produce short non-coding RNAs that are converted into double-stranded RNA by RDR2, and processed into 24-nt siRNA by DCL3. These siRNAs are loaded into AGO4/6 to target complementary non-coding transcripts generated by another plant-specific RNA polymerase (Pol V), and recruit the DNA methyltransferase DOMAINS REARRANGED METHYLASE 2 (DRM2) that catalyzes de novo DNA methylation, leading to TGS. This mechanism is known as RNA-directed DNA methylation (RdDM) [54–56]. In *Arabidopsis*, most TEs are maintained in a transcriptionally inactive state via cytosine methylation of their DNA sequences and chromatin compaction (mainly H3K9me2), thus limiting sequence accessibility to RNA polymerase II and transcription factors. These silenced TEs can nevertheless be used to inactivate active homologous TEs in the genome in trans via RdDM [57].

Once initiated, cytosine methylation is maintained by different methyltransferases depending on their sequence contexts. In *Arabidopsis*, METHYLTRANSFERASE 1 (MET1) maintains CG methylation (mCG), and CHROMOMETHYLASE 2 and 3 (CMT2/CMT3) maintain mCHH and mCHG levels, respectively. The activity of CMT2 and CMT3 is regulated by H3K9me2, as mCHH and mCHG recruit the histone H3K9 methyl transferases to dimethylate H3K9 residues, thus forming a self-reinforcing feedback loop between CHG/CHH methylation and H3K9me2 [54], and co-localization of both epigenetic marks. DNA methylation can be erased by DNA demethylases, such as REPRESSOR OF SILENCING 1 (ROS1) in *Arabidopsis*, whose transcription is tightly controlled by methylation monitoring sequences targeted both by RdDM and ROS1 itself, and acting as a genome-wide DNA methylation sensor.

As mentioned in the previous section, environmental stress (or developmental transitions) can activate TEs and release temporarily epigenetic silencing [43,51] (see Figure 3). However, the way these epigenetic marks are alleviated during such changes, and to what extent these epigenetic silencing mechanisms characterized in *Arabidopsis*, could be generalized to other plant species, and different TEs remain to be deciphered. LTRs of the stress inducible tobacco Tnt1 retrotransposon are targeted by 24-nt siRNAs, harbor a high CG/CHG and low CHH methylation level, but there is no evidence of a differential methylation associated with stress in this respect [58].

Another example is the *Arabidopsis* ONSEN Copia retrotransposon (ATCOPIA78), whose transcription is induced by heat-stress in the wild-type, but is not activated in plants treated with DNA methylation inhibitor 5-azacytidine, nor in *ddm1* (DECREASE IN DNA METHYLATION 1) plants with globally low levels of DNA methylation. This

indicates that a decrease in DNA methylation is not sufficient to immediately release *ONSEN* silencing [59], while in contrast, transcriptional and transposition activity of another *Arabidopsis* Copia LTR-RT (*ATCOPIA93* or *Evadé*, EVD) has been observed in DNA methylation mutants [60]. Accumulation of EVD copies or of several other TEs (including class II elements) have been reported in the F8 generation of different epiRILs (epigenetic recombinant inbred lines) populations, including the methylation-defective *met1* (for EVD) or *ddm1* mutants (for several TEs) crossed to wild-type accessions [60,61]. TE reactivation in these epiRILs seems to be progressive across several generations, despite the fact that DNA methylation pathways are restored after the first F1 cross. Does the first cross represent a type of (epi)genomic stress? Why do such reactivations occur in some epiRILs and not in others? Why are several generations needed before transposition occurrence? Recent studies started addressing these questions [62,63], but the mechanistic basis remains poorly understood and restricted to just a few examples in *Arabidopsis*.

Another recent work suggests that the methyltransferase CMT3 has an unexpected role in promoting *ONSEN* transcription by competing with CMT2, and thus preventing excessive CHH methylation marks and thus H3K9me2 in *ONSEN* sequences [64]. Taken together, the response of TEs following a given stress (or developmental signal) is complex. It may depend on a particular family or individual element, the presence of regulatory motifs in the TE sequence, or to the chromatin environment that allows the recruitment of regulatory factors to the promoter of the element [52]. For example, the linker histone H1 governs targeting of the nucleosome remodeler DDM1 to heterochromatic regions in order to provide access to DNA methyltransferases that participate in the RdDM pathway [65].

3.3. Transposable Element Biology

Unlike viruses, endogenous TEs must replicate in the germline in order to propagate within a population, and have to preserve the reproductive fitness of their host for their long-term maintenance. They need to develop self-control strategies to maintain a basal activity, in order to limit the deleterious effects of their transposition and avoid triggering epigenetic silencing. Copy number-related TE transcriptional silencing has been reported for different plant retrotransposons introduced into heterologous system, such as *Arabidopsis* plants transformed with autonomous tobacco Tto1 and Tnt1 LTR-RTs [66,67], with a transcriptional activity recorded at a threshold of four copies for Tnt1 and 2 for Tto1. Similarly, approximately 40 copies of the retrotransposon EVD seem to be required to achieve TGS in *Arabidopsis* [62]. Interestingly, active copies of EVD are preferentially transmitted through pollen [68], perhaps because this element is not expressed in this developmental stage [69], thus escaping epigenetic silencing.

Although not demonstrated in plants (to the best of our knowledge), plant TEs could also limit their activity by the production of inactive proteins interfering with active ones during the transposition cycle as shown for the yeast Ty1 retrotransposon, *Drosophila* P element, or the horn fly Himar1 mariner TIR transposon [70–72].

Another strategy to limit the deleterious effects of TE transposition is insertion specificity. TE distribution is the result of the balance between element integration site preference and selection [73]. In plant genomes, the TE genomic distribution differs depending on TE superfamilies or even family/lineages. For example, Ty3/Gypsy elements belonging to the CRM lineage/family (Table 1) and some LINEs retrotransposons are found mainly in centromeres (presumably due to the presence of a specific chromodomain linked to their integrase that probably interacts with the centromeric CENH3 protein and orientates their distribution). In contrast, other Ty3/Gypsy and most of Ty1/Copia LTR-RTs are distributed in blocks or scattered along chromosomes; *Brassica* MITEs are preferentially found in gene-rich regions [74,75]. Centromeric and pericentromeric regions, as well as knobs, are “TE islands” harboring high TE density with nested TEs and constitutive heterochromatin that is epigenetically maintained. Nesting of plant LTR-RTs in such regions is not random, as it targets preferentially older copies of the same family in the 3′-UTR, and seems to be sensitive to DNA sequence, secondary structure and chromatin environment [75], as

previously shown for other eukaryotes [73]. Recently, the role of the histone variant H2A.Z (incorporated preferentially in hypomethylated sequences and enriched in environmentally responsive genes in plants) in guiding integration of some Copia LTR-RTs has been evidenced in *Arabidopsis* [61,76].

3.4. Other Key-Players of Plant Genome Architecture

Besides the cytosine methylation and histone modification that have been already discussed, there are other biological features able to shape plant genomes and influence TE activity in plants. This includes the reproductive system (sexual/asexual), ability for hybridization and polyploidy (whole genome duplication), and the temporal dynamics of genome fractionation and the sequence elimination process through evolutionary time [77]. Genomic investigations conducted since 2000 revealed that seed plants experienced at least one round of whole genome duplication (WGD)/polyploidization [78]. Often considered as an 'evolutionary dead end', WGDs are now recognized as a short-term adaptive process, notably to stimulate genetic and epigenetic changes and rescue plant fitness under environmental stress, by means of DNA elimination, chromosome restructuring, or chromatin modifications [79,80].

One attractive and long-standing hypothesis is that TEs may play a central role in a situation of genomic shock triggered by hybridization/polyploidy [81]. Taking advantage of a disturbed context for epigenetic control, TEs may actively transpose and increase mutation rates or changes in gene regulation depending on the insertion sites. In such circumstances, TEs could also mediate genomic restructuring by inter-element recombination, chromosomal rearrangements and sequence elimination, as part of diploidization processes that return a polyploid genome to a diploid-like genome, such as sequence loss (duplicate genes, chromosomes or repetitive DNA), gene silencing, and altered chromosome pairing [82,83]. Analysis of TE transcription in wheat or *Arabidopsis* allopolyploid plants and their parental progenitors suggests that allopolyploidization can indeed activate some TE transcription [84,85], but this activation is not always followed by transposition and seems to be dependent on the TE family or plant species.

TE-related genomic changes have been thoroughly studied in several recent natural and synthetic allotetraploids of wheat, Brassica, Aegilops, or Nicotiana [84,86–88]. In these examples, no or minimal transposition events have been reported after genome merging and first generation, supporting the maintenance of active repressive mechanism of TE mobilization in the very first generation. By contrast, many TE-associated restructurings of parental insertions were observed as soon as F1 and shared across independent hybrids, as evidenced in Brassica, Aegilops or Nicotiana [87–89], suggesting non-random restructuring.

The study of different reciprocal plant hybrids/allopolyploids also evidenced that early genome changes often predominantly target one specific subgenome [88–90]. The underlying mechanisms and their timing through the polyploidization process are still not defined and may be species- or line-specific, but should involve epigenetic reprogramming, implying DNA/histone methylation, and siRNA and chromatin changes in the merged genomes [91]. If parental TE loads are divergent, the viability of the resulting hybrid will depend on the efficiency of the parental TE-derived siRNAs to control a TE population [92]. TEs can thus contribute to reproductive isolation and species formation, as recently demonstrated for the role of TE-derived small RNAs in pollen that trigger interploidy hybridization barriers in *Arabidopsis* [93,94]. Insertion profile comparison of TEs in three recent *Nicotiana* allopolyploids that arose from hybridization between increasingly divergent diploid parental species also show that quantitative imbalances between parental TE loads positively correlates with TE mobilization in the resulting allopolyploids [88]. This work highlighted TE quantitative imbalance among progenitors as an important factor for TE evolution in allopolyploids and supports the genome shock hypothesis proposed by McClintock.

4. Conclusions

Genomic TE content could be variable according to the species considered, and sometimes even to the genotype considered, such as for asian rice subspecies or maize lines (for a review, see [95]). TE composition and abundance results from a balance between counteracting forces, i.e., TE amplification and copy loss. TE dynamics are also impacted by host regulations and response to stress, and thus may reflect the evolution and adaptation of their host to the environment. Although potentially detrimental as a consequence of their transposition activity and repetitive nature, TEs evolved as sensors of external changes and drivers of plant evolution.

Comparison of eukaryotic genomes revealed that plants (angiosperms) have more plastic and less compartmentalized genomes than animals (mammals) [96,97]. This higher genome plasticity can be related to plant specific developmental features.

The first one is the alternation over generations of sporophytic (haploid) and gametophytic (diploid) states, coupled with double fertilization. During plant life cycles, genomes must be able to function in many different ploidy contexts—haploid (gametophyte), diploid (zygote), triploid (endosperm) contexts in diploid plants, and even higher ploidy levels in endoreduplicated cells or polyploid plants.

The second plant specificity is the absence of a sequestered germ line. As gametes differentiate late in plant development, hundreds of cell divisions take place between two zygote generations, offering many opportunities for mutations (as transpositions) to occur and to be transmitted to the offspring [96].

A third difference between plants and animals relies on the sessile nature of plants. In order to survive to stressful conditions, they might promote genomic restructuring and plasticity to generate genetic, biochemical and phenotypic novelties to adapt to external changes. By allowing transient TE silencing release during stress events, sessile plants developed a tool to generate moderate TE-mediated genomic changes which could be selected (or not) for plant adaptation to a changing environment.

Such a tightly controlled TE regulation is perhaps one of the reasons for the evolutive success of Angiosperms. Through epigenetic regulation, TEs are also thought to be involved in the maintenance of genome structure and processes, such as heterochromatin formation or reproductive isolation. Far from just being ‘junk DNA’, TEs are considered as a ‘blessing curse’ [98], and as major drivers of genome evolution and adaptation to environmental changes, even if only a part of their regulation mechanisms is understood. Each genome structure can also be considered as the result of a specific (unique) alternation of TE bursts and whole genome duplication (auto- or allopolyploidization), shaping accordingly silencing mechanisms and chromatin structure, and giving unique genomes [99,100]. Only the growing characterization of both TEs and epigenetic pathways in many plant species will help us to better understand the co-evolution of both partners.

Author Contributions: C.M. wrote the manuscript. C.M., F.B. and M.-A.G. revised and edited the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The IJPB benefits from the support of Saclay Plant Sciences (SPS) (ANR-17-EUR-0007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McClintock, B. Mutable Loci in Maize. *Annu. Rep. Dir. Dep. Genet. Carnegie Inst. Wash.* **1951**, *50*, 174–181.
2. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* **2009**, *326*, 1112–1115. [[CrossRef](#)] [[PubMed](#)]
3. Doolittle, W.F.; Sapienza, C. Selfish Genes, the Phenotype Paradigm and Genome Evolution. *Nature* **1980**, *284*, 601–603. [[CrossRef](#)] [[PubMed](#)]

4. Orgel, L.E.; Crick, F.H.C. Selfish DNA: The Ultimate Parasite. *Nature* **1980**, *284*, 604–607. [[CrossRef](#)]
5. Oliver, K.R.; Greene, W.K. Transposable Elements: Powerful Facilitators of Evolution. *BioEssays* **2009**, *31*, 703–714. [[CrossRef](#)]
6. Kazazian, H.H. Mobile Elements: Drivers of Genome Evolution. *Science* **2004**, *303*, 1626–1632. [[CrossRef](#)]
7. Biémont, C.; Vieira, C. Junk DNA as an Evolutionary Force. *Nature* **2006**, *443*, 521–524. [[CrossRef](#)]
8. Hua-Van, A.; Le Rouzic, A.; Boutin, T.S.; Filée, J.; Capy, P. The Struggle for Life of the Genome's Selfish Architects. *Biol. Direct* **2011**, *6*, 19. [[CrossRef](#)]
9. Lisch, D. How Important Are Transposons for Plant Evolution? *Nat. Rev. Genet.* **2013**, *14*, 49–61. [[CrossRef](#)]
10. Schulman, A.H.; Wicker, T. A Field Guide to Transposable Elements. In *Plant Transposons and Genome Dynamics in Evolution*; Fedoroff, N.V., Ed.; Wiley-Blackwell: Oxford, UK, 2013; pp. 15–40, ISBN 978-1-118-50015-6.
11. Bennetzen, J.L.; Wang, H. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Annu. Rev. Plant Biol.* **2014**, *65*, 505–530. [[CrossRef](#)]
12. Ibarra-Laclette, E.; Lyons, E.; Hernández-Guzmán, G.; Pérez-Torres, C.A.; Carretero-Paulet, L.; Chang, T.-H.; Lan, T.; Welch, A.J.; Juárez, M.J.A.; Simpson, J.; et al. Architecture and Evolution of a Minute Plant Genome. *Nature* **2013**, *498*, 94–98. [[CrossRef](#)] [[PubMed](#)]
13. Wicker, T.; Gundlach, H.; Spannagl, M.; Uauy, C.; Borrill, P.; Ramírez-González, R.H.; De Oliveira, R.; International Wheat Genome Sequencing Consortium; Mayer, K.F.X.; Paux, E.; et al. Impact of Transposable Elements on Genome Structure and Evolution in Bread Wheat. *Genome Biol.* **2018**, *19*, 103. [[CrossRef](#)] [[PubMed](#)]
14. Döring, H.P.; Tillmann, E.; Starlinger, P. DNA Sequence of the Maize Transposable Element Dissociation. *Nature* **1984**, *307*, 127–130. [[CrossRef](#)] [[PubMed](#)]
15. Pohlman, R.F.; Fedoroff, N.V.; Messing, J. The Nucleotide Sequence of the Maize Controlling Element Activator. *Cell* **1984**, *37*, 635–643. [[CrossRef](#)]
16. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A Unified Classification System for Eukaryotic Transposable Elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)] [[PubMed](#)]
17. Finnegan, D.J. Eukaryotic Transposable Elements and Genome Evolution. *Trends Genet.* **1989**, *5*, 103–107. [[CrossRef](#)]
18. Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten Things You Should Know about Transposable Elements. *Genome Biol.* **2018**, *19*, 199. [[CrossRef](#)]
19. Wells, J.N.; Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **2020**, *54*, 539–561. [[CrossRef](#)]
20. Neumann, P.; Novák, P.; Hošťáková, N.; Macas, J. Systematic Survey of Plant LTR-Retrotransposons Elucidates Phylogenetic Relationships of Their Polyprotein Domains and Provides a Reference for Element Classification. *Mob. DNA* **2019**, *10*, 1. [[CrossRef](#)]
21. Jiang, N.; Jordan, I.K.; Wessler, S.R. *Dasheng* and *RIRE2*. A Nonautonomous Long Terminal Repeat Element and Its Putative Autonomous Partner in the Rice Genome. *Plant Physiol.* **2002**, *130*, 1697–1705. [[CrossRef](#)]
22. Wicker, T.; Guyot, R.; Yahiaoui, N.; Keller, B. CACTA Transposons in Triticeae. A Diverse Family of High-Copy Repetitive Elements. *Plant Physiol.* **2003**, *132*, 52–63. [[CrossRef](#)] [[PubMed](#)]
23. Stritt, C.; Thieme, M.; Roulin, A.C. Rare Transposable Elements Challenge the Prevailing View of Transposition Dynamics in Plants. *Am. J. Bot.* **2021**, *108*, 1310–1314. [[CrossRef](#)] [[PubMed](#)]
24. Arkhipova, I.R. Distribution and Phylogeny of *Penelope*-Like Elements in Eukaryotes. *Syst. Biol.* **2006**, *55*, 875–885. [[CrossRef](#)] [[PubMed](#)]
25. Craig, R.J.; Yushenova, I.A.; Rodriguez, F.; Arkhipova, I.R. An Ancient Clade of *Penelope*-Like Retroelements with Permuted Domains Is Present in the Green Lineage and Protists, and Dominates Many Invertebrate Genomes. *Mol. Biol. Evol.* **2021**, *38*, 5005–5020. [[CrossRef](#)] [[PubMed](#)]
26. Lin, X.; Faridi, N.; Casola, C. An Ancient Transkingdom Horizontal Transfer of *Penelope*-like Retroelements from Arthropods to Conifers. *Genome Biol. Evol.* **2016**, *8*, 1252–1266. [[CrossRef](#)] [[PubMed](#)]
27. Heitkam, T.; Holtgräwe, D.; Dohm, J.C.; Minoche, A.E.; Himmelbauer, H.; Weisshaar, B.; Schmidt, T. Profiling of Extensively Diversified Plant LINES Reveals Distinct Plant-Specific Subclades. *Plant J.* **2014**, *79*, 385–397. [[CrossRef](#)] [[PubMed](#)]
28. Wenke, T.; Döbel, T.; Sörensen, T.R.; Junghans, H.; Weisshaar, B.; Schmidt, T. Targeted Identification of Short Interspersed Nuclear Element Families Shows Their Widespread Existence and Extreme Heterogeneity in Plant Genomes. *Plant Cell* **2011**, *23*, 3117–3128. [[CrossRef](#)]
29. Seibt, K.M.; Schmidt, T.; Heitkam, T. The Conserved 3' Angio-domain Defines a Superfamily of Short Interspersed Nuclear Elements (SINEs) in Higher Plants. *Plant J.* **2020**, *101*, 681–699. [[CrossRef](#)]
30. Bao, W.; Jurka, M.G.; Kapitonov, V.V.; Jurka, J. New Superfamilies of Eukaryotic DNA Transposons and Their Internal Divisions. *Mol. Biol. Evol.* **2009**, *26*, 983–993. [[CrossRef](#)]
31. De Assis, R.; Baba, V.Y.; Cintra, L.A.; Gonçalves, L.S.A.; Rodrigues, R.; Vanzela, A.L.L. Genome Relationships and LTR-Retrotransposon Diversity in Three Cultivated *Capsicum* L. (Solanaceae) Species. *BMC Genom.* **2020**, *21*, 237. [[CrossRef](#)]
32. Feschotte, C.; Pritham, E.J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* **2007**, *41*, 331–368. [[CrossRef](#)] [[PubMed](#)]
33. Kapitonov, V.V.; Jurka, J. Helitrons on a Roll: Eukaryotic Rolling-Circle Transposons. *Trends Genet.* **2007**, *23*, 521–529. [[CrossRef](#)] [[PubMed](#)]
34. Piednoël, M.; Gonçalves, I.R.; Higuët, D.; Bonnivard, E. Eukaryote DIRS1-like Retrotransposons: An Overview. *BMC Genom.* **2011**, *12*, 621. [[CrossRef](#)] [[PubMed](#)]

35. Wang, D.; Zheng, Z.; Li, Y.; Hu, H.; Wang, Z.; Du, X.; Zhang, S.; Zhu, M.; Dong, L.; Ren, G.; et al. Which Factors Contribute Most to Genome Size Variation within Angiosperms? *Ecol. Evol.* **2021**, *11*, 2660–2668. [[CrossRef](#)]
36. Piegu, B.; Guyot, R.; Picault, N.; Roulin, A.; Saniyal, A.; Kim, H.; Collura, K.; Brar, D.S.; Jackson, S.; Wing, R.A.; et al. Doubling Genome Size without Polyploidization: Dynamics of Retrotransposition-Driven Genomic Expansions in *Oryza Australiensis*, a Wild Relative of Rice. *Genome Res.* **2006**, *16*, 1262–1269. [[CrossRef](#)]
37. Macas, J.; Novák, P.; Pellicer, J.; Čížková, J.; Koblížková, A.; Neumann, P.; Fuková, I.; Doležel, J.; Kelly, L.J.; Leitch, I.J. In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe Fabeae. *PLoS ONE* **2015**, *10*, e0143424. [[CrossRef](#)]
38. Hawkins, J.S.; Kim, H.; Nason, J.D.; Wing, R.A.; Wendel, J.F. Differential Lineage-Specific Amplification of Transposable Elements Is Responsible for Genome Size Variation in *Gossypium*. *Genome Res.* **2006**, *16*, 1252–1261. [[CrossRef](#)]
39. Bertoli, D.J.; Cannon, S.B.; Froenicke, L.; Huang, G.; Farmer, A.D.; Cannon, E.K.S.; Liu, X.; Gao, D.; Clevenger, J.; Dash, S.; et al. The Genome Sequences of *Arachis Duranensis* and *Arachis Ipaensis*, the Diploid Ancestors of Cultivated Peanut. *Nat. Genet.* **2016**, *48*, 438–446. [[CrossRef](#)]
40. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory Activities of Transposable Elements: From Conflicts to Benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)]
41. Vicient, C.M. Transcriptional Activity of Transposable Elements in Maize. *BMC Genom.* **2010**, *11*, 601. [[CrossRef](#)]
42. Johns, M.A.; Mottinger, J.; Freeling, M. A Low Copy Number, Copia-like Transposon in Maize. *EMBO J.* **1985**, *4*, 1093–1101. [[CrossRef](#)] [[PubMed](#)]
43. Grandbastien, M.-A. LTR Retrotransposons, Handy Hitchhikers of Plant Regulation and Stress Response. *Biochim. Biophys. Acta BBA Gene Regul. Mech.* **2015**, *1849*, 403–416. [[CrossRef](#)] [[PubMed](#)]
44. Naito, K.; Zhang, F.; Tsukiyama, T.; Saito, H.; Hancock, C.N.; Richardson, A.O.; Okumoto, Y.; Tanisaka, T.; Wessler, S.R. Unexpected Consequences of a Sudden and Massive Transposon Amplification on Rice Gene Expression. *Nature* **2009**, *461*, 1130–1134. [[CrossRef](#)] [[PubMed](#)]
45. Orozco-Arias, S.; Isaza, G.; Guyot, R. Retrotransposons in Plant Genomes: Structure, Identification, and Classification through Bioinformatics and Machine Learning. *Int. J. Mol. Sci.* **2019**, *20*, 3837. [[CrossRef](#)]
46. Casacuberta, J.M.; Grandbastien, M.-A. Characterisation of LTR Sequences Involved in the Protoplast Specific Expression of the Tobacco Tnt1 Retrotransposon. *Nucleic Acids Res.* **1993**, *21*, 2087–2093. [[CrossRef](#)]
47. Takeda, S.; Sugimoto, K.; Otsuki, H.; Hirochika, H. 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.* **1999**, *18*, 383–393. [[CrossRef](#)]
48. Grandbastien, M.-A.; Audeon, C.; Bonnard, E.; Casacuberta, J.M.; Chalhoub, B.; Costa, A.-P.P.; Le, Q.H.; Melayah, D.; Petit, M.; Poncet, C.; et al. Stress Activation and Genomic Impact of Tnt1 Retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **2005**, *110*, 229–241. [[CrossRef](#)]
49. Galindo-González, L.; Mhiri, C.; Deyholos, M.K.; Grandbastien, M.-A. LTR-Retrotransposons in Plants: Engines of Evolution. *Gene* **2017**, *626*, 14–25. [[CrossRef](#)]
50. Beguiristain, T.; Grandbastien, M.-A.; Puigdomènech, P.; Casacuberta, J.M. Three Tnt1 Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco. Consequences for Retrotransposon Control and Evolution in Plants. *Plant Physiol.* **2001**, *127*, 212–221. [[CrossRef](#)]
51. Lisch, D. Epigenetic Regulation of Transposable Elements in Plants. *Annu. Rev. Plant Biol.* **2009**, *60*, 43–66. [[CrossRef](#)]
52. Lanciano, S.; Mirouze, M. Transposable Elements: All Mobile, All Different, Some Stress Responsive, Some Adaptive? *Curr. Opin. Genet. Dev.* **2018**, *49*, 106–114. [[CrossRef](#)] [[PubMed](#)]
53. Sigman, M.J.; Slotkin, R.K. The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *Plant Cell* **2016**, *28*, 304–313. [[CrossRef](#)] [[PubMed](#)]
54. Zhang, H.; Lang, Z.; Zhu, J.-K. Dynamics and Function of DNA Methylation in Plants. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 489–506. [[CrossRef](#)] [[PubMed](#)]
55. Borges, F.; Martienssen, R.A. The Expanding World of Small RNAs in Plants. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 727–741. [[CrossRef](#)] [[PubMed](#)]
56. Fultz, D.; Choudury, S.G.; Slotkin, R.K. Silencing of Active Transposable Elements in Plants. *Curr. Opin. Plant Biol.* **2015**, *27*, 67–76. [[CrossRef](#)] [[PubMed](#)]
57. Matzke, M.A.; Mosher, R.A. RNA-Directed DNA Methylation: An Epigenetic Pathway of Increasing Complexity. *Nat. Rev. Genet.* **2014**, *15*, 394–408. [[CrossRef](#)]
58. Hernández-Pinzón, I.; Cifuentes, M.; Hénaff, E.; Santiago, N.; Espinás, M.L.; Casacuberta, J.M. The Tnt1 Retrotransposon Escapes Silencing in Tobacco, Its Natural Host. *PLoS ONE* **2012**, *7*, e33816. [[CrossRef](#)]
59. Ito, H.; Gaubert, H.; Bucher, E.; Mirouze, M.; Vaillant, I.; Paszkowski, J. An siRNA Pathway Prevents Transgenerational Retrotransposition in Plants Subjected to Stress. *Nature* **2011**, *472*, 115–119. [[CrossRef](#)]
60. Mirouze, M.; Reinders, J.; Bucher, E.; Nishimura, T.; Schneeberger, K.; Ossowski, S.; Cao, J.; Weigel, D.; Paszkowski, J.; Mathieu, O. Selective Epigenetic Control of Retrotransposition in *Arabidopsis*. *Nature* **2009**, *461*, 427–430. [[CrossRef](#)]

61. Quadrana, L.; Etcheverry, M.; Gilly, A.; Caillieux, E.; Madoui, M.-A.; Guy, J.; Bortolini Silveira, A.; Engelen, S.; Baillet, V.; Wincker, P.; et al. Transposition Favors the Generation of Large Effect Mutations That May Facilitate Rapid Adaption. *Nat. Commun.* **2019**, *10*, 3421. [[CrossRef](#)]
62. Mari-Ordóñez, A.; Marchais, A.; Etcheverry, M.; Martin, A.; Colot, V.; Voinnet, O. Reconstructing de Novo Silencing of an Active Plant Retrotransposon. *Nat. Genet.* **2013**, *45*, 1029–1039. [[CrossRef](#)] [[PubMed](#)]
63. Rigal, M.; Becker, C.; Pélissier, T.; Pogorelnik, R.; Devos, J.; Ikeda, Y.; Weigel, D.; Mathieu, O. Epigenome Confrontation Triggers Immediate Reprogramming of DNA Methylation and Transposon Silencing in *Arabidopsis thaliana* F1 Epihybrids. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2083–E2092. [[CrossRef](#)] [[PubMed](#)]
64. Nozawa, K.; Chen, J.; Jiang, J.; Leichter, S.M.; Yamada, M.; Suzuki, T.; Liu, F.; Ito, H.; Zhong, X. DNA Methyltransferase CHROMOMETHYLASE3 Prevents ONSEN Transposon Silencing under Heat Stress. *PLoS Genet.* **2021**, *17*, e1009710. [[CrossRef](#)]
65. Zemach, A.; Kim, M.Y.; Hsieh, P.-H.; Coleman-Derr, D.; Eshed-Williams, L.; Thao, K.; Harmer, S.L.; Zilberman, D. The *Arabidopsis* Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell* **2013**, *153*, 193–205. [[CrossRef](#)] [[PubMed](#)]
66. Hirochika, H.; Okamoto, H.; Kakutani, T. Silencing of Retrotransposons in *Arabidopsis* and Reactivation by the *Ddm1* Mutation. *Plant Cell* **2000**, *12*, 357–368. [[CrossRef](#)] [[PubMed](#)]
67. Pérez-Hormaeche, J.; Potet, F.; Beauclair, L.; Le Masson, I.; Courtial, B.; Bouché, N.; Lucas, H. Invasion of the *Arabidopsis* Genome by the Tobacco Retrotransposon Tnt1 Is Controlled by Reversible Transcriptional Gene Silencing. *Plant Physiol.* **2008**, *147*, 1264–1278. [[CrossRef](#)]
68. Reinders, J.; Mirouze, M.; Nicolet, J.; Paszkowski, J. Parent-of-origin Control of Transgenerational Retrotransposon Proliferation in *Arabidopsis*. *EMBO Rep.* **2013**, *14*, 823–828. [[CrossRef](#)]
69. Lee, S.C.; Ernst, E.; Berube, B.; Borges, F.; Parent, J.-S.; Ledon, P.; Schorn, A.; Martienssen, R.A. *Arabidopsis* Retrotransposon Virus-like Particles and Their Regulation by Epigenetically Activated Small RNA. *Genome Res.* **2020**, *30*, 576–588. [[CrossRef](#)]
70. Saha, A.; Mitchell, J.A.; Nishida, Y.; Hildreth, J.E.; Ariberre, J.A.; Gilbert, W.V.; Garfinkel, D.J. A *Trans* -Dominant Form of Gag Restricts Ty1 Retrotransposition and Mediates Copy Number Control. *J. Virol.* **2015**, *89*, 3922–3938. [[CrossRef](#)]
71. Misra, S.; Rio, D.C. Cytotype Control of *Drosophila* P Element Transposition: The 66 Kd Protein Is a Repressor of Transposase Activity. *Cell* **1990**, *62*, 269–284. [[CrossRef](#)]
72. Lampe, D.J.; Akerley, B.J.; Rubin, E.J.; Mekalanos, J.J.; Robertson, H.M. Hyperactive Transposase Mutants of the Himar1 Mariner Transposon. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11428–11433. [[CrossRef](#)] [[PubMed](#)]
73. Sultana, T.; Zamborlini, A.; Cristofari, G.; Lesage, P. Integration Site Selection by Retroviruses and Transposable Elements in Eukaryotes. *Nat. Rev. Genet.* **2017**, *18*, 292–308. [[CrossRef](#)] [[PubMed](#)]
74. Li, S.-F.; Su, T.; Cheng, G.-Q.; Wang, B.-X.; Li, X.; Deng, C.-L.; Gao, W.-J. Chromosome Evolution in Connection with Repetitive Sequences and Epigenetics in Plants. *Genes* **2017**, *8*, 290. [[CrossRef](#)] [[PubMed](#)]
75. Jedlicka, P.; Lexa, M.; Vanat, I.; Hobza, R.; Kejnovsky, E. Nested Plant LTR Retrotransposons Target Specific Regions of Other Elements, While All LTR Retrotransposons Often Target Palindromes and Nucleosome-Occupied Regions: In Silico Study. *Mob. DNA* **2019**, *10*, 50. [[CrossRef](#)]
76. Roquis, D.; Robertson, M.; Yu, L.; Thieme, M.; Julkowska, M.; Bucher, E. Genomic Impact of Stress-Induced Transposable Element Mobility in *Arabidopsis*. *Nucleic Acids Res.* **2021**, *49*, 10431–10447. [[CrossRef](#)]
77. Wendel, J.F.; Lisch, D.; Hu, G.; Mason, A.S. The Long and Short of Doubling down: Polyploidy, Epigenetics, and the Temporal Dynamics of Genome Fractionation. *Curr. Opin. Genet. Dev.* **2018**, *49*, 1–7. [[CrossRef](#)]
78. Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, L.P.; Hu, Y.; Liang, H.; Soltis, P.S.; et al. Ancestral Polyploidy in Seed Plants and Angiosperms. *Nature* **2011**, *473*, 97–100. [[CrossRef](#)]
79. Alix, K.; Gérard, P.R.; Schwarzacher, T.; Heslop-Harrison, J.S. (Pat) Polyploidy and Interspecific Hybridization: Partners for Adaptation, Speciation and Evolution in Plants. *Ann. Bot.* **2017**, *120*, 183–194. [[CrossRef](#)]
80. Van de Peer, Y.; Mizrachi, E.; Marchal, K. The Evolutionary Significance of Polyploidy. *Nat. Rev. Genet.* **2017**, *18*, 411–424. [[CrossRef](#)]
81. McClintock, B. The Significance of Responses of the Genome to Challenge. *Science* **1984**, *226*, 792–801. [[CrossRef](#)]
82. Soltis, P.S.; Marchant, D.B.; Van de Peer, Y.; Soltis, D.E. Polyploidy and Genome Evolution in Plants. *Curr. Opin. Genet. Dev.* **2015**, *35*, 119–125. [[CrossRef](#)] [[PubMed](#)]
83. Vicent, C.M.; Casacuberta, J.M. Impact of Transposable Elements on Polyploid Plant Genomes. *Ann. Bot.* **2017**, *120*, 195–207. [[CrossRef](#)] [[PubMed](#)]
84. Kashkush, K.; Feldman, M.; Levy, A.A. Transcriptional Activation of Retrotransposons Alters the Expression of Adjacent Genes in Wheat. *Nat. Genet.* **2003**, *33*, 102–106. [[CrossRef](#)]
85. Madlung, A.; Tyagi, A.P.; Watson, B.; Jiang, H.; Kagochi, T.; Doerge, R.W.; Martienssen, R.; Comai, L. Genomic Changes in Synthetic *Arabidopsis* Polyploids: Genomic Changes in *Arabidopsis* Polyploids. *Plant J.* **2004**, *41*, 221–230. [[CrossRef](#)]
86. Sarilar, V.; Palacios, P.M.; Rousselet, A.; Ridet, C.; Falque, M.; Eber, F.; Chèvre, A.; Joets, J.; Brabant, P.; Alix, K. Allopolyploidy Has a Moderate Impact on Restructuring at Three Contrasting Transposable Element Insertion Sites in Resynthesized *Brassica napus* Allotetraploids. *New Phytol.* **2013**, *198*, 593–604. [[CrossRef](#)] [[PubMed](#)]
87. Senerchia, N.; Felber, F.; Parisod, C. Genome Reorganization in F₁ Hybrids Uncovers the Role of Retrotransposons in Reproductive Isolation. *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 20142874. [[CrossRef](#)]

88. Mhiri, C.; Parisod, C.; Daniel, J.; Petit, M.; Lim, K.Y.; Dorlhac de Borne, F.; Kovarik, A.; Leitch, A.R.; Grandbastien, M. Parental Transposable Element Loads Influence Their Dynamics in Young *Nicotiana* Hybrids and Allotetraploids. *New Phytol.* **2019**, *221*, 1619–1633. [[CrossRef](#)]
89. Lukens, L.N.; Pires, J.C.; Leon, E.; Vogelzang, R.; Oslach, L.; Osborn, T. Patterns of Sequence Loss and Cytosine Methylation within a Population of Newly Resynthesized *Brassica Napus* Allopolyploids. *Plant Physiol.* **2006**, *140*, 336–348. [[CrossRef](#)]
90. Koh, J.; Soltis, P.S.; Soltis, D.E. Homeolog Loss and Expression Changes in Natural Populations of the Recently and Repeatedly Formed Allotetraploid *Tragopogon mirus* (Asteraceae). *BMC Genom.* **2010**, *11*, 97. [[CrossRef](#)]
91. Parisod, C.; Alix, K.; Just, J.; Petit, M.; Sarilar, V.; Mhiri, C.; Ainouche, M.; Chalhoub, B.; Grandbastien, M.-A. Impact of Transposable Elements on the Organization and Function of Allopolyploid Genomes: Research Review. *New Phytol.* **2010**, *186*, 37–45. [[CrossRef](#)]
92. Martienssen, R.A. Heterochromatin, Small RNA and Post-Fertilization Dysgenesis in Allopolyploid and Interploid Hybrids of *Arabidopsis*: Research Review. *New Phytol.* **2010**, *186*, 46–53. [[CrossRef](#)] [[PubMed](#)]
93. Borges, F.; Parent, J.-S.; Van Ex, F.; Wolff, P.; Martínez, G.; Köhler, C.; Martienssen, R.A. Transposon-Derived Small RNAs Triggered by MiR845 Mediate Genome Dosage Response in *Arabidopsis*. *Nat. Genet.* **2018**, *50*, 186–192. [[CrossRef](#)] [[PubMed](#)]
94. Martinez, G.; Wolff, P.; Wang, Z.; Moreno-Romero, J.; Santos-González, J.; Conze, L.L.; DeFraia, C.; Slotkin, R.K.; Köhler, C. Paternal EasiRNAs Regulate Parental Genome Dosage in *Arabidopsis*. *Nat. Genet.* **2018**, *50*, 193–198. [[CrossRef](#)]
95. Vitte, C.; Fustier, M.-A.; Alix, K.; Tenaillon, M.I. The Bright Side of Transposons in Crop Evolution. *Brief. Funct. Genom.* **2014**, *13*, 276–295. [[CrossRef](#)]
96. Kejnovsky, E.; Leitch, I.J.; Leitch, A.R. Contrasting Evolutionary Dynamics between Angiosperm and Mammalian Genomes. *Trends Ecol. Evol.* **2009**, *24*, 572–582. [[CrossRef](#)]
97. Murat, F.; de Peer, Y.V.; Salse, J. Decoding Plant and Animal Genome Plasticity from Differential Paleo-Evolutionary Patterns and Processes. *Genome Biol. Evol.* **2012**, *4*, 917–928. [[CrossRef](#)]
98. Dubin, M.J.; Mittelsten Scheid, O.; Becker, C. Transposons: A Blessing Curse. *Curr. Opin. Plant Biol.* **2018**, *42*, 23–29. [[CrossRef](#)]
99. Springer, N.M.; Lisch, D.; Li, Q. Creating Order from Chaos: Epigenome Dynamics in Plants with Complex Genomes. *Plant Cell* **2016**, *28*, 314–325. [[CrossRef](#)]
100. Zhao, M.; Zhang, B.; Lisch, D.; Ma, J. Patterns and Consequences of Subgenome Differentiation Provide Insights into the Nature of Paleopolyploidy in Plants. *Plant Cell* **2017**, *29*, 2974–2994. [[CrossRef](#)]