

## Review Article

# On the Coevolution of Transposable Elements and Plant Genomes

Peter Cíváň,<sup>1</sup> Miroslav Švec,<sup>1</sup> and Pavol Hauptvogel<sup>2</sup>

<sup>1</sup> Department of Genetics, Faculty of Natural Sciences, Comenius University, Mlynská dolina B1, 842 15 Bratislava 4, Slovakia

<sup>2</sup> Plant Production Research Center, Gene Bank of Slovak Republik, Bratislavská cesta 122, 921 68 Piešťany, Slovakia

Correspondence should be addressed to Peter Cíváň, civan@fns.uniba.sk

Received 1 June 2011; Revised 23 August 2011; Accepted 6 September 2011

Academic Editor: Curtis C. Daehler

Copyright © 2011 Peter Cíváň et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Plant genomes are unique in an intriguing feature: the range of their size variation is unprecedented among living organisms. Although polyploidization contributes to this variability, transposable elements (TEs) seem to play the pivotal role. TEs, often considered intragenomic parasites, not only affect the genome size of the host, but also interact with other genes, disrupting and creating new functions and regulatory networks. Coevolution of plant genomes and TEs has led to tight regulation of TE activity, and growing evidence suggests their relationship became mutualistic. Although the expansions of TEs represent certain costs for the host genomes, they may also bring profits for populations, helping to overcome challenging environmental (biotic/abiotic stress) or genomic (hybridization and allopolyploidization) conditions. In this paper, we discuss the possibility that the possession of inducible TEs may provide a selective advantage for various plant populations.

## 1. Transpositional Strategies, Distribution, and Regulation of TEs

Transposable elements (TEs) comprise a palette of immensely diverse DNA structures that can be unified by the following definition: they all are (or have been) able to insert themselves (or new copies of themselves) into new locations within genome. According to their mechanism of transposition, TEs can be classified [1] into class I elements (retroelements) transposing through an RNA intermediate, and class II elements (DNA transposons) moving only via DNA. The major superfamilies of class I are *Ty1-copia* and *Ty3-gypsy* retrotransposons, while class II is represented by TIR (terminal inverted repeat) elements and *Helitrons*, which are sometimes classified separately [2]. Among both retrotransposons and transposons, nonautonomous forms (e.g., MITEs, SINEs, and LARDs) are quite prevalent [3], utilizing the transpositional machinery of autonomous TEs.

A significant portion of plant genomes is constituted by class I elements (specifically LTR retrotransposons, with direct long terminal repeats at both ends), which replicate in a “copy-and-paste” manner. In brief (according to [3, 4]), the genomic DNA copy of a retroelement is transcribed into mRNA that enters the cytosol, similarly to standard

DNA transcripts. The information of the mRNA is translated, typically creating a structural protein GAG and a polyprotein POL. These protein products associate with other retroelement mRNA copies and pack them into virus-like particles. Within these structures, dimerized mRNA copies are reversely transcribed into cDNA and the whole complex enters the nucleus, where the new cDNA copy integrates at a new site. This mode of transposition, if not suppressed, allows retroelements to massively increase their copy numbers, resulting in a rapid expansion of genome size.

Compared to retrotransposons, class II elements are anticipated to have a smaller potential of increasing their copy number. Owing to their “cut-and-paste” insertional mechanism mediated by transposases, multiplication arises only when a transposon from a recently replicated genomic region is transposed to a region about to undergo replication [5] or in cases when the excision site is repaired by gene conversion, using the sister chromatid as a template [4]. Nevertheless, short DNA transposons like MITEs (miniature inverted-repeat TEs) can be remarkably effective in increasing their copy numbers each generation [6, 7]. In the genome of *Lotus japonicus*, the copy number of detected MITEs is similar to, or higher than, the copy numbers of major LTR retrotransposon superfamilies [8], and the overall

contribution to the genome size is smaller only because of the short length of MITEs (200–500 bp).

For most of the TEs in plant genomes, a certain balance between TE proliferation and minimal damage to the host has evolved. This balance is widely achieved by epigenetic silencing, an important feature of which is its reversibility [9]. Epigenetic suppression of TEs are realized on different levels—from general transcriptional control by TE-unspecific histone modification and siRNA-directed DNA methylation [10–12], to posttranslational processes, like the species-specific control of nuclear localization of transposase [13, 14]. Silencing of TEs can be immensely effective. For example, LTR retrotransposons constitute a large part of the *Gossypium* genome, but almost no transcripts were found in the *Gossypium* EST database [15]. Interestingly, a variety of LINE-like transcripts have been found in the same EST libraries, suggesting different levels of suppression for particular TE classes. But attenuation of TEs is not irreversible. Disruption of epigenetic silencing patterns and consequent derepression and proliferation of TEs are thought to be associated with two natural phenomena—interspecific hybridization [16–18] and biotic/abiotic stress [19–23].

TEs are not evenly distributed along chromosomes. DNA transposons are overrepresented in gene-rich or euchromatic regions [7, 24], avoiding exons [7], while class I elements concentrate in gene-poor, heterochromatic regions around centromeres [24–28]. Exceptions to this general pattern are, for example, *copia* retroelements of maize overrepresented in euchromatin [29], or the FIDEL retrotransposon absent from heterochromatin in *Arachis* [30]. Authors frequently suppose the existence of yet unknown mechanisms of region-specific TE targeting. However, the uneven distributions of class I and class II elements across the genomes might be satisfactorily explained by probabilistic principles (as follows), without the need to hypothesize any targeting mechanism.

DNA transposons in heterochromatic regions may be less likely to be transposed because of the DNA topology inaccessible for transposases. On the contrary, DNA transposons in euchromatic or genic regions are being transposed frequently, and owing to the transposition strategy, the excised element is prone to reinsert at a nearby genomic location, which is likely to be euchromatic too (the shorter the distance between the original and novel position, the higher the probability that the two loci are in the same condition). The passive copies in the heterochromatin are eventually removed from the genome while only the active copies have chances to multiply, and thus colonize the euchromatin. Transpositions into exons are mostly filtered out by natural selection, so the majority of DNA transposons is observed in introns, or 5' and 3' adjacent regions. The less severe consequences of short TE insertions for splicing might be the reason why MITEs predominate in introns over other elements.

The situation is different for LTR retrotransposons. New copies of class I elements are generated outside the nucleus—in the cytosol, thus the probability for a new copy to reintegrate in the proximity of the maternal element is extremely low. Without any targeting mechanism, retroelements integrate randomly across the whole genome, and the observed uneven distribution can be attributed to

the following factors: (i) the majority of transpositions within or near genes is filtered out by selection because of the usually deleterious effects; and (ii) LTR retrotransposons (or TEs in general) are removed more efficiently from highly-recombining regions because the mechanisms of removal are recombination dependent [31–33]. This hypothesis is supported by the finding that *Sorghum* LTR retrotransposons younger than 10,000 years appear to be randomly distributed along chromosomes [34]. Distribution patterns of TEs are therefore likely to be a function of the transpositional strategy and age of the individual TE family, affected by methylation [35] and some genomic particularities of the host species (e.g., gene density and recombination landscapes).

## 2. Effects of Transposable Elements on Plant Genomes

**2.1. Genome Size.** The most obvious effect of TE existence in plants is their impact on the genome size. TE fraction can be as low as ~15% in small plant genomes, and as high as >70% in large plant genomes (Table 1) while the number of genes remains roughly constant [36] (of course, the latter does not hold true for polyploids). The correlation between the proportion of TEs—specifically LTR retrotransposons—and the physical length of the genome is so evident in the examined plant species [36, 37] that the genome size can be generally regarded as a linear function of TE content, and the dynamics of LTR retrotransposons as the major contributor to 1C-value differences among plants.

To answer the question what is the “typical” TE contribution to the length of plant genome, one needs some factual idea of a “typical” plant 1C-value. For this purpose, Tenaillon et al. [37] use an arithmetic mean of the 1C-values provided by Bennett and Leitch [38]. Their Plant DNA C-value database currently comprises 6,287 and 204 entries for angiosperms and gymnosperms, respectively; mostly based on Feulgen microdensitometry and flow cytometry data. Although this dataset probably cannot be considered a representative sample of all land plants (e.g., monocots being over-represented), as the most comprehensive one it still provides informative insights into plant genome size variation. The average genome size (1C) of angiosperms (flowering plants) and gymnosperms sampled in this database is 5.809 Gbp and 18.157 Gbp, respectively. However, the distribution of 1C-values, especially in angiosperms, shows strong positive skewness (Figure 1), with the data unevenly distributed around the mean (72.3% of the examined angiosperm species having smaller genomes than the mean value of the dataset). For such distributions, the median is a better indication of central tendency than the arithmetic mean. The median values of the angiosperm and gymnosperm genome size data sampled in the Plant DNA C-value database are 2.401 Gbp (1.416 Gbp for eudicots; 5.746 Gbp for monocots) and 17.506 Gbp, respectively. In other words, half of the examined flowering plants has genome sizes below 2.4 Gbp. Surprisingly, the interval 0.4–0.6 Gbp represents here the modal category for angiosperm 1C-values, comprising 8.38% of examined species (Figure 1(b)). Hence, *Oryza sativa*, *Lotus japonicus*, *Medicago truncatula*, *Vitis*

TABLE 1: Known TE proportions in plant genomes when various authors provide different values, the range is given. Inconsistent estimates of TE content result from incomplete genomic representation and/or varying bioinformatical approaches. Examination of complete genomes and application of the same TE discovery pipeline is therefore essential for comparative analyses on intra- and interspecies level. No data on TE content is available for gymnosperms, and angiosperms with large genomes (>2.4 Gbp) are underrepresented. In most genomes, *Helitrons* were not surveyed appropriately.

Species	Predominant fertilization	Genome size (Gbp)	TE content (%)	DNA transposon content (%)	Retrotransposon content (%)	Genomic fraction examined	References
<i>Arabidopsis thaliana</i>	S	0.125	6–15.36	2–3	4	Complete/draft genome	[12, 25, 32]
<i>Arabidopsis lyrata</i>	C	0.207	23.48	ns	ns	Complete/draft genome	[12, 73]
<i>Brachypodium distachyon</i>	S	0.272	26.17	4.77	21.4	Complete/draft genome	[74]
<i>Carica papaya</i>	C	0.372	51.9	ns	ns	Complete/draft genome	[75]
<i>Cucumis melo</i>	C	0.454	8.13	0.93	7.2	BAC clones (6.7 Mbp)	[76]
<i>Oryza sativa</i>	S	0.472	35	ns	17	Complete chr. 10; complete/draft genome	[24, 39]
<i>Lotus japonicus</i>	S	0.472	30.8	0.97–8.1	10.4–19.23	TAC clones (32.4 Mbp); 2/3 genome	[8, 77]
<i>Medicago truncatula</i>	S	0.475	38	ns	9.6*	BAC clones (0.233 Mbp)	[40, 78]
<i>Populus trichocarpa</i>	C	0.485	42	2.5**	10.3**	Complete/draft genome	[79]
<i>Vitis vinifera</i>	S	0.487–0.505	21.5–41.4	1.98	20	Complete/draft genome	[41, 80, 81]
<i>Brassica oleracea</i>	C	0.6	20	6	14	Complete/draft genome	[32]
<i>Sorghum bicolor</i>	S	0.73	62.5–59	7.5/17	55/42	Complete/draft genome	[34, 82]
<i>Glycine max</i>	S	1.1	58	16	42	Complete/draft genome (3.42 Mbp)	[38, 83]
<i>Gossypium herbaceum</i>	S	1.667	52.1	<0.1	52	Small insert genomic library (3.42 Mbp)	[84]
<i>Zea mays</i>	C	2.3	85	8.6	75	Complete/draft genome	[29]
<i>Aegilops tauschii</i>	S	4.988	68.2	13.3	53.5	Small insert genomic library (2.911 Mbp)	[38, 85]
<i>Hordeum vulgare</i>	S	5.7	88.1	ns	ns	BAC clones (0.44 Mbp)	[86]
<i>Secale cereale</i>	C	8.093	69.3	5	64.3	BAC clones (2.033 Mbp)	[38, 87]

S—self-pollination; C—cross-pollination; ns—not specified; \*only full-length LTR retrotransposons analysed; \*\*only TEs with homology to known repeat elements considered.

*vinifera*, *Brassica oleracea*, and *Populus trichocarpa* have “typical” genome lengths (Table 1), therefore it can be implied that transposable elements typically constitute one-third of angiosperm genomes.

Genome size changes are not unidirectional. TE fragments and solo LTRs have been found to constitute a significant fraction of repetitive elements (e.g., [39–43]), and

are believed to be remnants of removed TEs [31]. There is evidence that transpositional bursts can be followed by DNA loss [44], and it has been reported that the removal of LTR retrotransposons can proceed with different efficiency in distinct species [33, 40, 42]. However, it should be noted that the estimations of a retroelements’ half-life assume constant removal rates for repetitive sequences, and rely on the

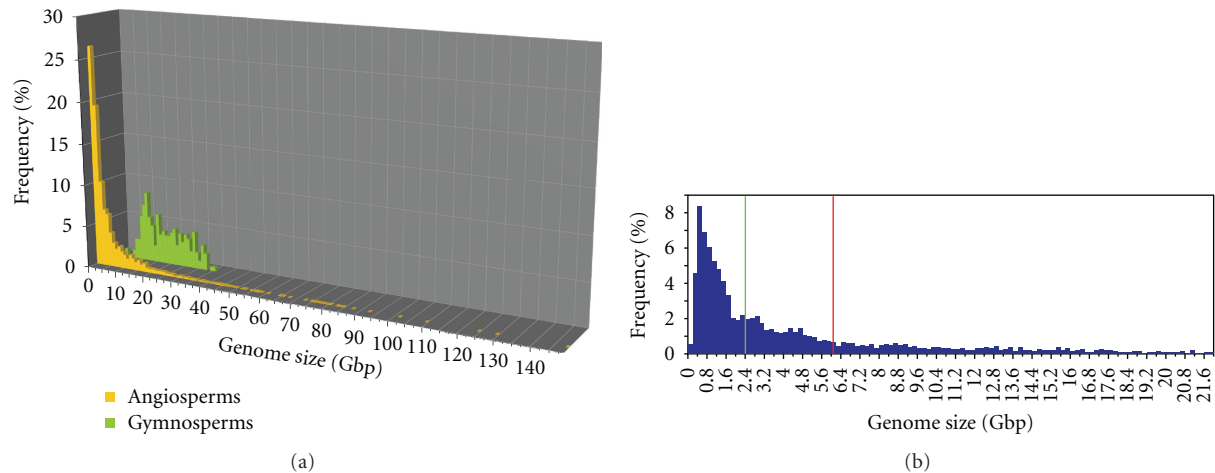


FIGURE 1: Frequency distribution for *spermatophyta* 1C-values. The dataset consists of all angiosperm (6,287) and gymnosperm (204) entries in the Plant DNA C-value database [38] (retrieved 27/05/2011). (a) Comparison of angiosperm and gymnosperm 1C-value frequency distribution. Genomes of *spermatophyta* were split into categories with 1 Gbp intervals (X-axis); the proportion (percent) of each category (vertical axis) is shown separately for the two groups. Interestingly, angiosperms (mean = 5.809 Gbp,  $s_x$  = 9.5287, median = 2.401 Gbp) and gymnosperms (mean = 18.157 Gbp,  $s_x$  = 7.4496, median = 17.506 Gbp) do not display similar distributions. Despite the extreme range of 1C-values, 95% of flowering plants fall within 0–22 Gbp. The range of gymnosperm 1C-values is smaller; however, the overall genome size is higher, with 95% of species falling within 7–33 Gbp. (b) Higher resolution of angiosperm 1C frequency distribution. X-axis shows genome size categories with 0.2 Gbp increments and the vertical axis indicates the proportion of each category (percent). Mean and median values are indicated by the red and green line, respectively. The mean of the first half of the histogram (left to the median) is 1.049 Gbp; the mean of the second half is 10x larger (not indicated). Hence, the average “small” (<2.4 Gbp) and “large” (>2.4 Gbp) angiosperm genomes are ~1 Gbp and ~10 Gbp long, respectively. Modal interval (0.4–0.6 Gbp) includes 8.38% of flowering plants.

molecular clock principle; therefore, the revealed interspecies differences in TE survival should be interpreted cautiously.

Disregarding the mutational effects of TEs at this point, a dramatic increase of noncoding or repetitive fraction in the genome theoretically raises the nutritional and time requirements for DNA replication and maintenance in each cell (leading to putative costs formulated as the “large genome constraint hypothesis,” [45]). The higher nutritional and time demands may lead to decreased fecundity and prolonged generation time, the two main constituents affecting selective advantage [46]. While such changes are apparently disfavoring in populations of unicellular organisms, the growing genome size does not seem to impose an evident and unambiguous selective disadvantage in *spermatophyta* (seed plants) [45, 47].

Offering one intriguing example: phosphorus (P) is known to be the limiting nutrient in most soils (reviewed in [48]). As it is one of the basic components of nucleic acids, plants with significantly larger genomes have higher P demands for DNA synthesis, leaving lesser amounts for other essential cell processes (e.g., ATP or phospholipid dependent). Therefore, it seems natural to anticipate that plants with large C-values should loose the ecological competition against the plants whose genomes are several-fold smaller. Despite that, genome sizes in examined land plants range 2056-fold [49], with *Genlisea margaretae* (1C = 63.4 Mbp = 0.0648 pg) [50] and *Trillium hageae* (1C = 132.5 pg) [51] on the opposite poles. Relatively large, almost 22-fold differences in 1C-values have also been reported for members of a single genus (*Eleocharis*) at the same ploidy level [52].

If we reject the unlikely possibility that such enormous genome size variation in plants is attributable to stochastic processes only, the question arises as to why some plants maintain relatively small genomes (Figure 1) while others sink into genomic obesity [53]. In theory, there are two possible causes that may allow transposable elements, and thus genomes to expand extensively: (i) deficiency in the mechanisms of suppression and/or removal of TEs from the genome and (ii) selective advantage that favours individuals and/or populations with high TE activity.

In relation to the former possibility (i), Weil and Martienssen [9] compare the interaction of TEs and host genomes to resistance to pathogens, and hypothesize that as transposons evolve ways around host silencing, host organisms evolve new genes for silencing, perhaps through duplication and subfunctionalization. This idea is supported by the discovery of positive selection acting on some LTR retrotransposons in the rice genome [54]. On rare occasions, mutant variants of retroelements can escape host recognition and rapidly amplify, leading to what is commonly observed as bursts of amplification. Consequently, the detected variability in TE activity in time and taxa can be attributed to different phases of the “host-parasite” interaction. The occasional escape of TEs from the host suppression via random mutations and positive selection seems plausible; however, such events should exhibit roughly the same periodicity in all genomes, assuming similar substitutional rates of repetitive DNA among species. Therefore, this “red queen race” in itself can hardly explain the enormous differences in plant genome sizes. Moreover, the host-pathogen analogy also implies that the species with below-average performance in TE

suppression are evolutionary disadvantaged to those species who have successfully prevented transpositional bursts of parasitic retroelements. Comparing the plant species with small and large genomes, we lack any direct evidence for such generalization.

The above-mentioned conflicts direct the attention to the latter alternative (ii), which suggests plant genome size variation to be caused by a differential selective advantage of TE possession acting in distinct species or populations (i.e., the presence and activity of transposable elements might be beneficial in some species/populations while detrimental in others). Testability of this hypothesis depends on identifying the nature of such advantage; therefore, other effects of TEs on plant genomes need to be carefully considered (see below).

**2.2. Mutability.** The discovery of transposable elements was accompanied by the observation of their mutability [55], which unlike the growing genome size, can often have a phenotypic manifestation. Transpositions into the coding regions of genes are usually deleterious; however, those transpositional events that passed through the sieve of selection can induce a variety of genetic changes, including interrupting host genes, creating different expression forms, changing intron length, and affecting expression levels of adjacent genes [56]. Downregulation of genes may be caused not only by TE insertions disrupting the promoters, but more likely by siRNA-guided DNA methylation, which is primarily directed to suppress the TE activity but affects the expression of nearby genes too [35]. Whole-genome differences in TE-siRNA interactions have such dramatic effects on expressional patterns that they may contribute to speciation [12]. Among recently reported TE-induced mutations are cluster-shaped somatic variation in grapevine caused by the insertion of *Hatvine1-rrm* DNA transposon in the *VvTFL1A* gene promoter [57]; flower color gene mutation caused by *TgmExpress1* transposition into the intron 2 of *F3H* gene in soybean [58]; or transposon-induced DNA methylation of *CmWIP1* promoter leading to sex determination in melon [59].

Comprehensive and genome-wide analyses of TE mutability have been accomplished on the fully sequenced genome of rice. According to [39], while LTR retrotransposons constitute ~17% of the rice genome, 22% of these sequences lie within putative or established rice genes. Within the genic regions, fragmented elements have been predominantly identified, and full-length elements are rare [39]. Available genomic sequences of the two rice subspecies—*japonica* and *indica*—provide a powerful resource for comparative and functional genomic analyses, which has been utilized by Huang et al. [56] to study transposon insertion polymorphisms (TIPs). Interestingly, more than 10% of TIPs between Nipponbare and 93-11 rice cultivars were located in expressed gene regions. Roughly half of those TIPs occurred in introns, often resulting in alternative splicing, and more than a third were found in [−1, −250] regions, relative to the transcription start site. Effects of TE insertions within the promoters are particularly impressive in the case of two genes, causing 18-fold upregulation and 23-fold

downregulation of gene expression [56]. In another study [7], high-throughput sequencing was utilized to determine 1,664 insertion sites of *mPing* transposon in a population of 24 rice plants. Subsequent comparative microarray analysis concluded that the vast majority of TE insertions either have no impact, or preferentially enhance transcription under normal conditions. However, seven out of ten loci, unaffected by *mPing* insertion under normal conditions, were inducible by salt and cold stress. Scanning *mPing* sequences for cis-acting plant regulatory elements resulted in identification of 96 putative regulatory motifs, one-third of which were stress responsive. These experiments demonstrate that the *mPing* transposon, resembling a mobile gene enhancer, provides new binding sites for transcription factors or other regulatory proteins, and may actually benefit the host by creating potentially useful allelic variants and novel, stress-inducible regulatory networks [7].

Among the most intriguing features of transposable elements is the ability of certain classes to capture gene fragments. The potential to contribute to gene evolution by combining genes, exons, and introns into novel functional units is most apparent in *Helitrons*. Although these elements were initially challenging to identify due to the absence of typical TE structural features [60, 61], an effective structure-based program has been developed recently, leading to the detection of thousands *Helitrons* in several plant genomes [61]. The frequency of the gene capture is particularly striking in the genome of maize [62, 63]. For example, Morgante et al. [62] randomly selected nine genic insertions polymorphic in maize inbred lines and demonstrated that eight of them are nonautonomous *Helitrons*, each containing between one and seven different fragments of host genes. Yang and Bennetzen [63] have shown on a genome-wide scale that 60% of maize *Helitrons* contains captured fragments of nuclear genes, 4% of which are under purifying selection, and another 4% exhibiting apparent adaptive selection, which suggests beneficial effects for the host or *Helitron* transposition/retention. Although the vast majority of the genes captured by *Helitrons* are incomplete, defective copies of conserved functional genes (including exons and also introns), a fraction of those gene fragments may serve as the template for interfering RNAs or new gene functions via exon shuffling, expanding the repertoire of mutable changes provided by TEs.

The last significant mutable effect of TEs to be mentioned is the macrotransposition, that is, a transposition involving two physically close, interacting elements, and an intervening chromosomal segment. Such transposon pairs may produce other complex rearrangements, including deletions, inversions, and reshuffling of the intertransposon segment [64], thus macrotransposition can be another contributor to genome divergence and speciation.

### 3. From Individuals to Populations: From Junk to Treasure

In the study of coevolution of TEs and plant genomes, two different populations affected by inconsistent selection forces should be recognized—the population of transposable

elements within the genomic niche of an individual, and the population of diverse individuals within an ecological niche. In this section, the term population refers to the latter one.

Since the early work of Barbara McClintock, the presence and action of transposable elements in the host genomes has been studied from the perspective of an individual. Because the mutational consequences of TE activity on genes and their expression patterns (see above) are undirected by the host and principally random, it has become apparent that the uncontrolled TE transposition or expansion is, in the vast majority of cases neutral, detrimental, or lethal for an individual. The absence of evident benefits of the possession of transposable elements for an individual led to TEs being regarded as an archetype of selfish or parasitic DNA, whose only functional aim is to reproduce itself, regardless of the effect on the host genome. The large genomic regions occupied by ancient, suppressed, or still active transposable elements have acquired a label “junk DNA”—an unnecessary burden for cell and organism.

The individual-based viewpoint in the TE research was needed, because a fundamental description of TE diversity, prevalence, and *modus operandi* was required, and also understandable, because the technical possibilities to study TE dynamics on the population level were unavailable until recently. However, evolution acts on populations, not on individuals. Some recent studies have drawn the attention to population processes related to TEs [37, 65–67], and it is becoming clearer that to answer the questions about the origin, evolution, function, and importance of transposable elements in plant genomes, it is necessary to move the research focus from individuals to populations.

Environmental stresses are known initiators of TE activity [20–23, 68], and diverse effects of transpositional events on the expression of adjacent genes have been reported [7, 35, 56, 57, 59, 68]. Although the mutational impact of TE bursts is likely to be detrimental for an individual, TE activity creates new variability in the population, providing raw material for selection forces. An illustrative example of how effective the TEs can be in generating genetic diversity is provided by the activity of the *mPing* DNA transposon in some rice strains. With roughly 40 new transpositions of *mPing* per plant per generation, even small populations contain thousands of new insertions, a large portion of which upregulates genes in their vicinity under stress conditions [6, 7].

Hence, TE activity may actually help the population to overcome changing environmental conditions and adapt to new ecological settings. Under diversifying selection, this ability of quick adaptations is likely to outbalance the costs of decreased fitness of some individuals, or the possible large genome constraint. From this perspective, the escape of TEs from the silenced state resembles more a regulated response to cope with stress on population level rather than an undesired side effect of stress exposure, an idea initially hypothesized by McClintock [55] as the “response to genomic shock.” Possession of a mechanism that can boost the evolutionary changes and be switched on and off depending on the situation might be the decisive factor for the survival or extinction of a population in changing environments. It is suggestive to hypothesize that

transposable elements might represent such a tool and were actually “invented,” or at least modified by eukaryotes to fulfil this function. And if plants possess and use an autonomous mechanism to control TE proliferation, it means they also have a basic control over their own genome size.

The hypothesis of transposable elements as intrinsic tools for increasing genetic variability has some testable implications (i)–(iii). For example, (i) TE-driven stimulation of variability could be especially beneficial for species, populations, or genomic regions exposed to strong diversifying selection (e.g., host-pathogen systems). Such entities would be therefore expected to possess more transposable elements and their TE dynamics to be more responsive to stress conditions. Interestingly, Nielen et al. [30] have found LTR retrotransposon FIDEL associated with conserved *Arachis* genes less frequently than what was expected by chance, but its presence close to fast-evolving NBS genes (resistance gene analogues) was in agreement with random distribution.

(ii) Asexually reproducing organisms and self-pollinating plants, which lack the opportunity to recombine their genetic material, might profit from the enhanced diversity sustained by the TEs, as suggested for rice by Naito et al. [7]. However, the outcrosser *Arabidopsis lyrata* shows 2–3 times higher TE content than the selfer *A. thaliana* (Table 1). The relationship between the mating system and TE dynamics for these two relatives has been studied [68, 69], and among the possible causes are differences in effective population size and related stochastic processes, or more deleterious consequences of the accumulation of recessive mutations in self-pollinating plants. Self-pollinators might also be more efficient in suppressing TEs owing to more rapid fixation of epigenetic silencing patterns.

On the other hand, Bestor [70] assumes that the aggressiveness of transposons in self-fertilizing sexuals is self-limited comparing to outcrossing sexuals, where the transposon fixation is nearly certain provided that the coefficient of selection imposed by the transposon is less than 0.5 when there is one or more transposition events per generation. It implies that self-pollinators are not expected to have higher transposon content than related outcrossers, unless the TEs provide some net benefits to the host. While *Triticeae* seems to be an interesting tribe for such comparisons (Table 1), more comprehensive data on the TE content is required.

(iii) Taxa adapted to environmentally stable niches, such as ocean depths or high altitudes, would be expected to contain significantly less TEs comparing to the organisms from more unstable environments, which supposedly overcame multiple bursts of TE activity. An excellent opportunity to study these expectations is provided by three diploid sunflower species. *Helianthus anomalus*, *H. deserticola*, and *H. paradoxus* are independently derived via hybridization events between the same two parental taxa, *H. annuus* and *H. petiolaris*. The three hybrid taxa encountered a rapid, retrotransposon-mediated genome expansion [71, 72], and all of them occupy habitats considered abiotically extreme relative to either parental species. *H. anomalus* and *H. deserticola* inhabit arid desert-like environments whereas *H. paradoxus* occurs exclusively in saline environments.

Interestingly, the scale of copy number increase for *copia* LTR retrotransposons differs considerably among the three sunflower hybrid species, with a 3.7-fold increase in copy number in the genome of *H. paradoxus* (relative to the average parental species value) versus a lower 1.7-fold and 2.2-fold increase for *H. anomalus* and *H. deserticola*, respectively [72]. The copy number increase of *gypsy* LTR retrotransposons is even more stunning, with 5.6- to 23.6-fold multiplication in the hybrid taxa, compared to parental populations (which did not differ significantly) [71]. Ungerer et al. [71] suggest that hybridization, abiotic stress, or both may have been involved in this extensive retrotransposon proliferation. In either case, it is tempting to hypothesize that this is an example of TE proliferation being “switched on” by the host regulatory mechanisms (possibly coevolving with the TEs) as an action to elicit mutational consequences potentially helpful in adapting to new environments.

#### 4. Conclusion

Genome-wide and population-based examinations of similar projections have the potential to illuminate the role of transposable elements in speciation, adaptation, and in the evolution of plant genomes, in general. Besides, a detailed knowledge of TE activity regulation is required to understand the coevolution of TEs and plant genomes, and the extent of consequential benefits utilizable by plants. During the past decades, the notion of TEs has varied from autotelic junk to valued tool of evolutionary response. In the light of new evidence, the terms like “junk” or “selfish” DNA, and even “host genome” and “defence mechanisms for TE suppression” are becoming more misleading than ever. At present, comparative and functional genomic studies targeted on TE population dynamics and TE-cell interactions, supported by high-throughput technologies, are on the way to finalize this paradigm shift.

#### Acknowledgments

The authors would like to thank two anonymous reviewers for helpful suggestions and constructive critique. This paper was supported by the Slovak Research and Development Agency under the Contracts no. APVV-0770-07 and no. APVV-0661-10.

#### References

- [1] T. Wicker, F. Sabot, A. Hua-Van et al., “A unified classification system for eukaryotic transposable elements,” *Nature Reviews Genetics*, vol. 8, no. 12, pp. 973–982, 2007.
- [2] V. V. Kapitonov and J. Jurka, “*Helitrons* on a roll: eukaryotic rolling-circle transposons,” *Trends in Genetics*, vol. 23, no. 10, pp. 521–529, 2007.
- [3] F. Sabot and A. H. Schulman, “Genomics of transposable elements in the Triticeae,” in *Genetics and Genomics of the Triticeae*, C. Feuillet and G. J. Muehlbauer, Eds., pp. 387–405, 2009.
- [4] M. Labrador and V. G. Corces, “Transposable element-host interactions: regulation of insertion and excision,” *Annual Review of Genetics*, vol. 31, pp. 381–404, 1997.
- [5] P. J. Russell, *iGenetics*, Benjamin/Cummings, San Francisco, Calif, USA, 1st edition, 2002.
- [6] K. Naito, E. Cho, G. Yang et al., “Dramatic amplification of a rice transposable element during recent domestication,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17620–17625, 2006.
- [7] K. Naito, F. Zhang, T. Tsukiyama et al., “Unexpected consequences of a sudden and massive transposon amplification on rice gene expression,” *Nature*, vol. 461, no. 7267, pp. 1130–1134, 2009.
- [8] S. Sato, Y. Nakamura, T. Kaneko et al., “Genome structure of the legume, *Lotus japonicus*,” *DNA Research*, vol. 15, no. 4, pp. 227–239, 2008.
- [9] C. Weil and R. Martienssen, “Epigenetic interactions between transposons and genes: lessons from plants,” *Current Opinion in Genetics and Development*, vol. 18, no. 2, pp. 188–192, 2008.
- [10] R. K. Tran, D. Zilberman, C. de Bustos et al., “Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*,” *Genome Biology*, vol. 6, no. 11, article R90, 2005.
- [11] F. J. Qin, Q. W. Sun, L. M. Huang, X. S. Chen, and D. X. Zhou, “Rice SUVH histone methyltransferase genes display specific functions in chromatin modification and retrotransposon repression,” *Molecular Plant*, vol. 3, no. 4, pp. 773–782, 2010.
- [12] J. D. Hollister, L. M. Smith, Y. L. Guo, F. Ott, D. Weigel, and B. S. Gaut, “Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 6, pp. 2322–2327, 2011.
- [13] T. Uchiyama, Y. Saito, H. Kuwabara et al., “Multiple regulatory mechanisms influence the activity of the transposon, Tam3, of *Antirrhinum*,” *New Phytologist*, vol. 179, no. 2, pp. 343–355, 2008.
- [14] K. Fujino, S. N. Hashida, T. Ogawa et al., “Temperature controls nuclear import of Tam3 transposase in *Antirrhinum*,” *Plant Journal*, vol. 65, no. 1, pp. 146–155, 2011.
- [15] G. Hu, J. S. Hawkins, C. E. Grover, and J. F. Wendel, “The history and disposition of transposable elements in polyploid *Gossypium*,” *Genome*, vol. 53, no. 8, pp. 599–607, 2010.
- [16] B. Liu and J. F. Wendel, “Retrotransposon activation followed by rapid repression in introgressed rice plants,” *Genome*, vol. 43, no. 5, pp. 874–880, 2000.
- [17] M. Petit, C. Guidat, J. Daniel et al., “Mobilization of retrotransposons in synthetic allotetraploid tobacco,” *New Phytologist*, vol. 186, no. 1, pp. 135–147, 2010.
- [18] X. Shan, Z. Liu, Z. Dong et al., “Mobilization of the active MITE transposons *mPing* and *Pong* in rice by introgression from wild rice (*Zizania latifolia* Griseb.),” *Molecular Biology and Evolution*, vol. 22, no. 4, pp. 976–990, 2005.
- [19] H. Ito, H. Gaubert, E. Bucher, M. Mirouze, I. Vaillant, and J. Paszkowski, “An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress,” *Nature*, vol. 472, no. 7341, pp. 115–120, 2011.
- [20] S. Pouteau, M. A. Grandbastien, and M. Boccar, “Microbial elicitors of plant defense responses activate transcription of a retrotransposon,” *Plant Journal*, vol. 5, no. 4, pp. 535–542, 1994.
- [21] S. R. Wessler, “Plant retrotransposons: turned on by stress,” *Current Biology*, vol. 6, no. 8, pp. 959–961, 1996.
- [22] M. A. Grandbastien, “Activation of plant retrotransposons under stress conditions,” *Trends in Plant Science*, vol. 3, no. 5, pp. 181–187, 1998.

- [23] P. Capy, G. Gasperi, C. Biéumont, and C. Bazin, "Stress and transposable elements: co-evolution or useful parasites?" *Heredity*, vol. 85, no. 2, pp. 101–106, 2000.
- [24] T. Sasaki, "The map-based sequence of the rice genome," *Nature*, vol. 436, no. 7052, pp. 793–800, 2005.
- [25] S. Kaul, H. L. Koo, J. Jenkins et al., "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*," *Nature*, vol. 408, no. 6814, pp. 796–815, 2000.
- [26] B. D. Peterson-Burch, D. Nettleton, and D. F. Voytas, "Genomic neighborhoods for *Arabidopsis* retrotransposons: a role for targeted integration in the distribution of the Metaviridae," *Genome Biology*, vol. 5, no. 10, article R78, 2004.
- [27] J. Du, Z. Tian, N. J. Bowen, J. Schmutz, R. C. Shoemaker, and J. Ma, "Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean," *Plant Cell*, vol. 22, no. 1, pp. 48–61, 2010.
- [28] C. Staginnus, C. Desel, T. Schmidt, and G. Kahl, "Assembling a puzzle of dispersed retrotransposable sequences in the genome of chickpea (*Cicer arietinum* L.)," *Genome*, vol. 53, no. 12, pp. 1090–1102, 2010.
- [29] P. S. Schnable, D. Ware, R. S. Fulton et al., "The B73 maize genome: complexity, diversity, and dynamics," *Science*, vol. 326, no. 5956, pp. 1112–1115, 2009.
- [30] S. Nielen, F. Campos-Fonseca, S. Leal-Bertioli et al., "FIDEL—a retrovirus-like retrotransposon and its distinct evolutionary histories in the A- and B-genome components of cultivated peanut," *Chromosome Research*, vol. 18, no. 2, pp. 227–246, 2010.
- [31] K. M. Devos, J. K. M. Brown, and J. L. Bennetzen, "Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*," *Genome Research*, vol. 12, no. 7, pp. 1075–1079, 2002.
- [32] X. Zhang and S. R. Wessler, "Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 15, pp. 5589–5594, 2004.
- [33] C. Vitte and J. L. Bennetzen, "Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17638–17643, 2006.
- [34] A. H. Paterson, J. E. Bowers, R. Bruggmann et al., "The *Sorghum bicolor* genome and the diversification of grasses," *Nature*, vol. 457, no. 7229, pp. 551–556, 2009.
- [35] J. D. Hollister and B. S. Gaut, "Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression," *Genome Research*, vol. 19, no. 8, pp. 1419–1428, 2009.
- [36] K. M. Devos, "Grass genome organization and evolution," *Current Opinion in Plant Biology*, vol. 13, no. 2, pp. 139–145, 2010.
- [37] M. I. Tenaillon, J. D. Hollister, and B. S. Gaut, "A triptych of the evolution of plant transposable elements," *Trends in Plant Science*, vol. 15, no. 8, pp. 471–478, 2010.
- [38] M. D. Bennett and I. J. Leitch, "Plant DNA C-values Database," release 5.0, December 2010, <http://data.kew.org/cvalues/>.
- [39] L. Gao, E. M. McCarthy, E. W. Ganko, and J. F. McDonald, "Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences," *BMC Genomics*, vol. 5, article 18, 2004.
- [40] H. Wang and J.-S. Liu, "LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in rice," *BMC Genomics*, vol. 9, article 382, 2008.
- [41] A. Benjak, A. Forneck, and J. M. Casacuberta, "Genome-wide analysis of the "cut-and-paste" transposons of grapevine," *PLoS ONE*, vol. 3, no. 9, Article ID e3107, 2008.
- [42] T. Wicker and 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 Research*, vol. 17, no. 7, pp. 1072–1081, 2007.
- [43] M. Charles, H. Belcram, J. Just et al., "Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat," *Genetics*, vol. 180, no. 2, pp. 1071–1086, 2008.
- [44] C. Vitte, O. Panaud, and H. Quesneville, "LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss," *BMC Genomics*, vol. 8, article 218, 2007.
- [45] C. A. Knight, N. A. Molinari, and D. A. Petrov, "The large genome constraint hypothesis: evolution, ecology and phenotype," *Annals of Botany*, vol. 95, no. 1, pp. 177–190, 2005.
- [46] L. M. Wahl and C. S. DeHaan, "Fixation probability favors increased fecundity over reduced generation time," *Genetics*, vol. 168, no. 2, pp. 1009–1018, 2004.
- [47] C. A. Knight and J. M. Beaulieu, "Genome size scaling through phenotype space," *Annals of Botany*, vol. 101, no. 6, pp. 759–766, 2008.
- [48] C. P. Vance, C. Uhde-Stone, and D. L. Allan, "Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource," *New Phytologist*, vol. 157, no. 3, pp. 423–447, 2003.
- [49] I. J. Leitch, J. M. Beaulieu, M. W. Chase, A. R. Leitch, and M. F. Fay, "Genome size dynamics and evolution in monocots," *Journal of Botany*, vol. 2010, Article ID 862516, 2010.
- [50] J. Greilhuber, T. Borsch, K. Müller, A. Worberg, S. Porembski, and W. Barthlott, "Smallest angiosperm genomes found in *Lentibulariaceae*, with chromosomes of bacterial size," *Plant Biology*, vol. 8, no. 6, pp. 770–777, 2006.
- [51] B. J. M. Zonneveld, "New record holders for maximum genome size in eudicots and monocots," *Journal of Botany*, vol. 2010, Article ID 527357, 2010.
- [52] F. Zedek, J. Šmerda, P. Šmarda, and P. Bureš, "Correlated evolution of LTR retrotransposons and genome size in the genus *eleocharis*," *BMC Plant Biology*, vol. 10, article 265, 2010.
- [53] J. L. Bennetzen and E. A. Kellogg, "Do plants have a one-way ticket to genomic obesity?" *Plant Cell*, vol. 9, no. 9, pp. 1509–1514, 1997.
- [54] R. S. Baucom, J. C. Estill, J. Leebens-Mack, and J. L. Bennetzen, "Natural selection on gene function drives the evolution of LTR retrotransposon families in the rice genome," *Genome Research*, vol. 19, no. 2, pp. 243–254, 2009.
- [55] B. McClintock, "The significance of responses of the genome to challenge," *Science*, vol. 226, no. 4676, pp. 792–801, 1984.
- [56] X. Huang, G. Lu, Q. Zhao, X. Liu, and B. Han, "Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice," *Plant Physiology*, vol. 148, no. 1, pp. 25–40, 2008.
- [57] L. Fernandez, L. Torregrosa, V. Segura, A. Bouquet, and J. M. Martinez-Zapater, "Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine," *Plant Journal*, vol. 61, no. 4, pp. 545–557, 2010.

- [58] G. Zabala and L. Vodkin, "Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in *Glycine max*," *BMC Plant Biology*, vol. 7, article 38, 2007.
- [59] A. Martin, C. Troadec, A. Boualem et al., "A transposon-induced epigenetic change leads to sex determination in melon," *Nature*, vol. 461, no. 7267, pp. 1135–1138, 2009.
- [60] C. Du, J. Caronna, L. He, and H. K. Dooner, "Computational prediction and molecular confirmation of Helitron transposons in the maize genome," *BMC Genomics*, vol. 9, article 51, 2008.
- [61] L. Yang and J. L. Bennetzen, "Structure-based discovery and description of plant and animal Helitrons," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 31, pp. 12832–12837, 2009.
- [62] M. Morgante, S. Brunner, G. Pea, K. Fengler, A. Zuccolo, and A. Rafalski, "Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize," *Nature Genetics*, vol. 37, no. 9, pp. 997–1002, 2005.
- [63] L. Yang and J. L. Bennetzen, "Distribution, diversity, evolution, and survival of Helitrons in the maize genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 47, pp. 19922–19927, 2009.
- [64] J. T. Huang and H. K. Dooner, "Macrotransposition and other complex chromosomal restructuring in maize by closely linked transposons in direct orientation," *Plant Cell*, vol. 20, no. 8, pp. 2019–2032, 2008.
- [65] C. E. Grover and J. F. Wendel, "Recent insights into mechanisms of genome size change in plants," *Journal of Botany*, vol. 2010, Article ID 382732, 2010.
- [66] A. Le Rouzic and G. Deceliere, "Models of the population genetics of transposable elements," *Genetical Research*, vol. 85, no. 3, pp. 171–181, 2005.
- [67] A. Le Rouzic, T. S. Boutin, and P. Capi, "Long-term evolution of transposable elements," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 49, pp. 19375–19380, 2007.
- [68] S. I. Wright, Q. H. Le, D. J. Schoen, and T. E. Bureau, "Population dynamics of an Ac-like transposable element in self- and cross-pollinating arabidopsis," *Genetics*, vol. 158, no. 3, pp. 1279–1288, 2001.
- [69] S. Lockton and B. S. Gaut, "The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*," *BMC Evolutionary Biology*, vol. 10, no. 1, article 10, 2010.
- [70] T. H. Bestor, "Sex brings transposons and genomes into conflict," *Genetica*, vol. 107, no. 1–3, pp. 289–295, 1999.
- [71] M. C. Ungerer, S. C. Strakosh, and Y. Zhen, "Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation," *Current Biology*, vol. 16, no. 20, pp. R872–R873, 2006.
- [72] T. Kawakami, S. C. Strakosh, Y. Zhen, and M. C. Ungerer, "Different scales of *Ty1/copia*-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species," *Heredity*, vol. 104, no. 4, pp. 341–350, 2010.
- [73] T. T. Hu, P. Pattyn, E. G. Bakker et al., "The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change," *Nature Genetics*, vol. 43, no. 5, pp. 476–483, 2011.
- [74] J. P. Vogel, D. F. Garvin, T. C. Mockler et al., "Genome sequencing and analysis of the model grass *Brachypodium distachyon*," *Nature*, vol. 463, no. 7282, pp. 763–768, 2010.
- [75] R. Ming, S. Hou, Y. Feng et al., "The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus)," *Nature*, vol. 452, no. 7190, pp. 991–996, 2008.
- [76] V. M. González, A. Benjak, E. M. Hénaff et al., "Sequencing of 6.7 Mb of the melon genome using a BAC pooling strategy," *BMC Plant Biology*, vol. 10, article 246, 2010.
- [77] D. Holligan, X. Zhang, N. Jiang, E. J. Pritham, and S. R. Wessler, "The transposable element landscape of the model legume *Lotus japonicus*," *Genetics*, vol. 174, no. 4, pp. 2215–2228, 2006.
- [78] S. B. Cannon, L. Sterck, S. Rombauts et al., "Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 40, pp. 14959–14964, 2006.
- [79] G. A. Tuskan, S. DiFazio, S. Jansson et al., "The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)," *Science*, vol. 313, no. 5793, pp. 1596–1604, 2006.
- [80] R. Velasco, A. Zharkikh, M. Troggio et al., "A high quality draft consensus sequence of the genome of a heterozygous grapevine variety," *PLoS ONE*, vol. 2, no. 12, Article ID e1326, 2007.
- [81] O. Jaillon, J. M. Aury, B. Noel et al., "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla," *Nature*, vol. 449, no. 7161, pp. 463–467, 2007.
- [82] J. Schmutz, S. B. Cannon, J. Schlueter et al., "Genome sequence of the palaeopolyploid soybean," *Nature*, vol. 463, no. 7278, pp. 178–183, 2010.
- [83] J. Du, D. Grant, Z. Tian et al., "SoyTEDb: a comprehensive database of transposable elements in the soybean genome," *BMC Genomics*, vol. 11, no. 1, article 113, 2010.
- [84] J. S. Hawkins, H. Kim, J. D. Nason, R. A. Wing, and J. F. Wendel, "Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*," *Genome Research*, vol. 16, no. 10, pp. 1252–1261, 2006.
- [85] W. Li, P. Zhang, J. P. Fellers, B. Friebe, and B. S. Gill, "Sequence composition, organization, and evolution of the core Triticeae genome," *Plant Journal*, vol. 40, no. 4, pp. 500–511, 2004.
- [86] T. Wicker, W. Zimmermann, D. Perovic et al., "A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley *Hv-elf4E* locus: recombination, rearrangements and repeats," *Plant Journal*, vol. 41, no. 2, pp. 184–194, 2005.
- [87] J. Bartoš, E. Paux, R. Kofler et al., "A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R," *BMC Plant Biology*, vol. 8, article 95, 2008.

