# The history and disposition of transposable elements in polyploid *Gossypium*

**Guanjing Hu, Jennifer S. Hawkins, Corrinne E. Grover, and Jonathan F. Wendel**

**Abstract:** Transposable elements (TEs) are a major component of plant genomes. It is of particular interest to explore the potential activation of TE proliferation, especially in hybrids and polyploids, which often are associated with rapid genomic and epigenetic restructuring. Here we explore the consequences of genomic merger and doubling on *copia* and *gypsy*-like *Gorge3* long terminal repeat (LTR) retrotransposons as well as on non-LTR long interspersed nuclear elements (LINEs) in allotetraploid cotton, *Gossypium hirsutum.* Using phylogenetic and quantitative methods, we describe the composition and genomic origin of TEs in polyploid *Gossypium.* In addition, we present information on ancient and recent transposition activities of the three TE types and demonstrate the absence of an impressive proliferation of TEs following polyploidization in *Gossypium.* Further, we provide evidence for present-day transcription of LINEs, a relatively minor component of *Gossypium* genomes, whereas the more abundant LTR retrotransposons display limited expression and only under stressed conditions.

*Key words: copia*, *gypsy*, LINE, genome evolution, repetitive DNA.

**Résumé :** Les éléments transposables (ET) constituent une part importante des génomes des plantes. L'activation potentielle de la prolifération des ET présente un intérêt particulier, spécialement dans le contexte des hybrides et des polyploïdes chez lesquels se produisent fréquemment des changements génomiques et épigénétiques rapides. Ici, les auteurs explorent les conséquences de la fusion et du doublement des génomes sur les rétrotransposons à LTR *copia* et *Gorge3* (de type *gypsy*), de même que sur les éléments LINE sans LTR chez le cotonnier allotétraploïde, *Gossypium hirsutum.* À l'aide de méthodes phylogénétiques et quantitatives, les auteurs décrivent la composition et l'origine génomique des ET chez le *Gossypium* polyploïde. De plus, les auteurs présentent des renseignements sur les activités de transposition, à la fois ancienne et récente, pour ces trois types d'ET et ils démontrent l'absence de prolifération majeure des ET suite à la polyploïdisation chez le *Gossypium*. Finalement, les auteurs apportent des évidences que la transcription des éléments LINE, une composante mineure des génomes du *Gossypium*, a encore lieu aujourd'hui tandis que les rétrotransposons à LTR, beaucoup plus abondants, présentent une expression limitée et seulement sous des conditions de stress.

*Mots-clés : copia*, *gypsy*, LINE, évolution génomique, ADN répété.

[Traduit par la Rédaction]

## Introduction

Transposable elements (TEs) are ubiquitous among plants, often comprising more than 50% of plant genomes. The vast majority of these elements are class I retrotransposons, which replicate through reverse transcription of RNA intermediates and subsequent integration into the host genome. This "copy-and-paste" mode of transposition allows retrotransposons to massively increase their copy numbers, resulting in a rapid expansion of genome size, as seen in maize, rice, cotton, and other plants (Hawkins et al. 2006; Morse et al. 2009; Piegu et al. 2006; SanMiguel and Bennetzen 1998; SanMiguel et al. 1998; Ungerer et al. 2006). Retrotransposons are further divided into two subfamilies, long terminal repeat (LTR) and non-LTR retrotransposons, based upon the presence of LTR sequences. The LTR retrotransposon subfamily, which includes Ty1-*copia* and Ty3-*gypsy* elements, represents the most widespread and abundant type of TEs in plants. The differential success of these elements among plant species has been recognized as a major influence in plant genome size variation, together with genome-wide duplications through polyploidy (Grover et al. 2008; Hawkins et al. 2006, 2008, 2009; Piegu et al. 2006). The non-LTR retrotransposon subfamily, which includes long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), is less represented in plants (Hawkins et al. 2006; Noma et al. 1999; Le et al. 2000; Turcotte et al. 2001), in contrast to its prominence in mammalian genomes (Lander et al. 2001; Venter et al. 2001).

In addition to affecting the size of a plant genome, active TEs can be highly mutagenic when insertions happen within or near genic regions. As such, it is generally held that the

Received 1 February 2010. Accepted 8 May 2010. Published on the NRC Research Press Web site at genome.nrc.ca on 17 July 2010.

Corresponding Editor: T. Bureau.

**G. Hu, C.E. Grover, and J.F. Wendel.**[1] Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA.
**J.S. Hawkins.** Department of Genetics, University of Georgia, Athens, GA 30602, USA.

[1]Corresponding author (e-mail: jfw@iastate.edu).

transposition of resident TEs should be controlled by the host genome so that the elements are quiescent; however, active elements are often described, some with the demonstrated ability to alter the structure and expression of "neighbor" genes. The resulting genetic novelty can sometimes facilitate the adaptation and evolution of the host genome (Hori et al. 2007; Jiang et al. 2004; Lai et al. 2005; Xiao et al. 2008; Yan et al. 2006; Zabala and Vodkin 2005). Moreover, transposons themselves serve as indispensable regulators of processes as diverse as heterochromatin formation (Zofall and Grewal 2006), telomere and centromere function (Neumann et al. 2007; Pidoux and Allshire 2005; Zofall and Grewal 2006), and development (Jones-Rhoades et al. 2006; Slotkin et al. 2009).

To better understand the impact and evolutionary significance of TEs on their host genomes, the accumulation and historical rate of transposition of TEs have been studied in many plants. These studies have revealed the dynamic nature of TEs among genomes and the punctuated nature of their amplification (Hawkins et al. 2006, 2008; Piegu et al. 2006; Wicker and Keller 2007). As TEs are largely quiescent in all genomes examined thus far, possible triggers of transpositional bursts have been identified, specifically environmental and genomic stresses (Capy et al. 2000; Wessler 1996) and especially interspecific hybridization (de Boer et al. 2007; Kashkush et al. 2002, 2003; Labrador et al. 1999; Liu and Wendel 2000; O'Neill et al. 1998; Petit et al. 2007, 2010; Petrov et al. 1995; Shan et al. 2005; Ungerer et al. 2006). This phenomenon is often interpreted as reflecting a generalized disruption of epigenetic suppression of TE activity caused by the merger of two diverged regulatory systems.

Because polyploidy is common in plants, and because it often entails the union of two diverged genomes, it is of interest to explore the possibility of TE activation in polyploids. Here we describe this effort for *Gossypium hirsutum*, the most important of the cultivated species of cotton. *Gossypium* is a monophyletic group composed of approximately 45 diploid species divided into 8 genome groups (A–G, K) and 5 polyploid species (AD). Since the origin of the genus, the diploid genome groups have evolved genome sizes that vary more than 3-fold, largely because of differential rates of TE proliferation and DNA removal among lineages (Hawkins et al. 2006, 2009). Approximately 1–2 million years ago (mya), an A-genome diploid species and a D-genome diploid species reunited in a common nucleus after experiencing independent and differential accumulation of TEs since their diversification 5–10 mya. This hybridization and subsequent polyploidization brought together two different complements of transposable elements, previously described by Hawkins et al. (2006), creating the potential for release from suppression or activation of TEs owing to "genomic shock" resulting from hybridization or polyploidization. To evaluate the fate of TEs in the polyploid *Gossypium* AD genome, we describe the TE composition of polyploid cotton as well as the genome of origin (A or D) and timing of transposition (pre- or post-polyploidization) of these TEs. In addition, we present information on the expression and possible mobility of these elements in polyploid *Gossypium*.

## Materials and methods

### Plant materials and phylogenetic analysis

Total genomic DNA was extracted from fresh young leaves of *G. hirsutum* 'Acala Maxxa' using the DNeasy Plant Mini Kit (QIAGEN Inc., Valencia, California). The reverse transcriptase (RT) regions of *gypsy*-like *Gorge3*, *copia*, and LINEs were amplified and sequenced using degenerate primers and conditions described in Hawkins et al. (2008, 2009). PCR products were purified using the QIAquick Gel Extraction Kit (QIAGEN Inc.), cloned with the pGEM -T Easy Vector System (Promega, Madison, Wisconsin), and sequenced using the Applied Biosystems 3730*xl* DNA Analyzer at the Iowa State University DNA facility. A total of 273 sequences were generated (*Gorge3*, 94; *copia*, 112; LINE, 67), aligned with MUSCLE (Edgar 2004), and subsequently inspected manually (sequence data have been submitted to GenBank under accession Nos. HM626729–HM627001). Neighbor-joining analysis was conducted on the aligned sequences using PAUP* (Swofford 2001). Uncorrected ("p") DNA/RNA distances were set for distance analysis and missing data were ignored for affected pairwise comparisons.

### Analysis of transposition events and TE composition in polyploid *Gossypium*

Estimation of the timing of recent TE transposition events for polyploid *Gossypium* was performed in the same manner as described in Hawkins et al. (2008, 2009), where the same element types (*gypsy*-like *Gorge3*, *copia*, and LINE) were amplified and characterized for 3 diploid *Gossypium* species, *G. herbaceum* (A genome), *G. raimondii* (D genome), and *G. exiguum* (K genome), and a phylogenetic outgroup, *Gossypioides kirkii*. Briefly, monophyletic TE clades with less than 10% average nucleotide divergence were considered "lineage-specific" (originating 5–7 mya), whereas clades with sequence divergence between 10% and 20% were considered "*Gossypium*-specific" (originating 10–12 mya). These predetermined limits, based on an understanding of the phylogenetic history and timing of divergence of the species involved, permit the timing of transposition events to be estimated using TE sequence divergence superimposed on both element and organism phylogenies. Pairwise nucleotide diversity ($\pi$) was calculated among all daughter sequences of monophyletic clades, and then the density of these values was plotted to obtain the frequency of nucleotide divergences.

In the polyploid AD genome, PCR-amplified TE sequences could have been derived from either the A or the D genome and could have originated as either "*Gossypium*-specific" or "lineage-specific" sequences from those genomes, the latter originating from either pre- or post-polyploidization amplification events. By assuming that the sequence divergence of TE orthologs between the A and D genomes of the polyploid is approximately equal to the divergence between those found in the diploid A and D genomes, as is the case for other sequences (Senchina et al. 2003), AD-genome TEs were considered "lineage-specific" of diploid genome origin when they clustered in only A- or D-genome clades with an average divergence of <10% and accordingly were assigned a parental origin. Cal-

culation of nucleotide diversity and clustering of AD-genome TEs with diploid orthologs was performed in R with the base package and with the phylogenetics package APE (Paradis et al. 2004).

### Survey of TE transcriptional activity

To explore for "active" elements in AD-genome *Gossypium*, we used BLASTN (Altschul et al. 1997) and the PCR sequences of *Gorge3*, *copia*, and LINE RT regions to search the 153 969 published *Gossypium* ESTs (Udall et al. 2006) and an additional 1.3 million recently generated EST reads from 454 sequencing (J. Udall et al., unpublished data). Matched *G. hirsutum* ESTs with an *E* value of <E–10 were considered to represent transcripts of potentially active TEs. Growing conditions that might correlate with TE transcription, such as biotic or abiotic stress, were determined from the EST library of origin and were noted.

## Results

### PCR amplification and phylogenetic analysis

Transposable elements represent a major component of *Gossypium* genomes (Hawkins et al. 2006). In previous work, we characterized the proliferation history of three types of TEs (the *gypsy*-like *Gorge3*, *copia*, and LINE elements) in diploid *Gossypium* by phylogenetic analysis of their PCR-amplified RT regions (Hawkins et al. 2008, 2009). Here, using the same sets of degenerate primers, we amplified the same RT regions of the three classes of TEs described above from the polyploid *G. hirsutum* (AD genome). A total of 273 sequences were generated: 94 *Gorge3*, 112 *copia*, and 67 LINE sequences.

Divergence analyses of these sequences were generally congruent with the nucleotide diversities based on the survey of 3 diploid genomes (A genome, D genome, and K genome) and the outgroup reference genome, *Gossypioides kirkii* (Table 1). LINE elements displayed greater nucleotide divergence inter se than did *Gorge3* and *copia* elements, consistent with previous findings in *Gossypium* (Hawkins et al. 2008, 2009; Senchina et al. 2003) and also with those in other plants which indicate that LINEs tend to be more diverse than LTR retrotransposons (Holligan et al. 2006; Schmidt et al. 1995; Schwarz-Sommer et al. 1987; Wright et al. 1996). For sequences from the parental diploids (A genome and D genome) and the AD genome, the range of nucleotide diversity values was narrowest for LINE sequences (0.38 to 0.41), intermediate for *Gorge3* elements (0.18 to 0.30), and greatest for *copia* elements (0.057 in the A genome, 0.271 in the D genome, and 0.363 in the AD genome), which suggests a more diverse *copia* population in the polyploid genome.

### *Gorge3* evolution in polyploid *Gossypium*

Ninety-four *Gorge3* sequences of 480 bp from the AD genome were aligned with 724 sequences from the diploid genomes (A, 235; D, 114; K, 211; and *Gossypioides kirkii*, 164) and subjected to neighbor-joining analysis (Fig. 1A). Congruent with previous research (Hawkins et al. 2008), most RT sequences from the diploids clustered into 4 major genome-specific clades and 2 "cotton-specific" clades that contain sequences from all diploid *Gossypium* species. As expected based on organismal history, *Gorge3* sequences from allopolyploid cotton were closely related to and interdigitated with sequences from the parental diploid A and D genomes. To gain insight into the origin of specific sequences or groups of sequences from the allopolyploid AD genome, we assessed nucleotide diversity in clades that include sequences from the AD allopolyploid and one parental genome to assign a probable genome of origin to each AD-genome sequence. Based on a previously described average of 10% sequence divergence between orthologous TEs from the A and D genomes (Grover et al. 2007; Hawkins et al. 2008), AD sequences in monophyletic clades composed solely of AD- and A-genome sequences that averaged <10% nucleotide divergence were considered to be *Gorge3* elements having an A-genome origin, whereas *Gorge3* sequences originating from the D genome were identified in monophyletic AD–D clades with <10% average nucleotide diversity. Notably, this analysis revealed 72 *Gorge3* sequences of A-genome origin yet only 2 of D-genome origin, suggesting either a strongly biased composition of genome-specific *Gorge3* elements in polyploid *Gossypium* or an amplification bias (discussed further below).

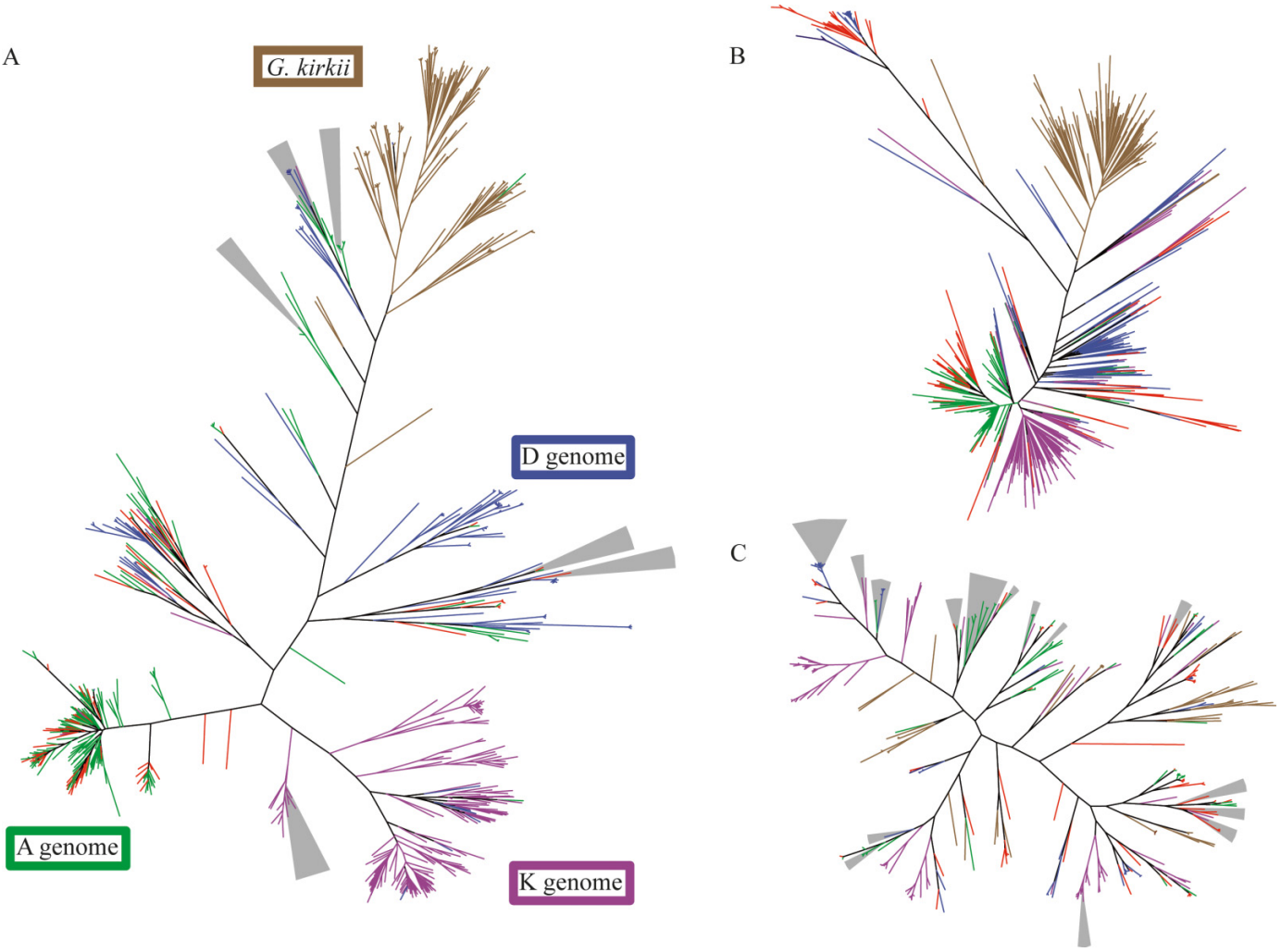
To detect and characterize recent amplification of *Gorge3* in polyploid *Gossypium*, we extracted the pairwise nucleotide divergences from AD-genome sequences and a "synthetic AD genome", the latter comprising A- and D-genome *Gorge3* sequences, and graphed these divergences on a divergence/time axis using previously published time curves for diploid cotton (Fig. 2A). The curve representing the synthetic AD genome largely mirrors the additive profile of the A and D lineage-specific transposition curves at about 2.5–3.5 and <1 mya; however, an extra peak at a sequence diversity of 0.07 was observed. This extra peak can be attributed to the inclusion of clades with <10% average nucleotide diversity that contain both A- and D-genome sequences, which were excluded from the previous separate analysis of each diploid genome. When viewed with the "synthetic AD genome" and the parental diploids, *Gorge3* in the AD genome exhibits little evidence of proliferation around the time of polyploid formation (ca. 1–2 mya). A burst at a sequence diversity of 0.05, which represents a time point prior to polyploid formation, is consistent with a burst of transposition in the A genome and mostly reflects the vertical inheritance of *Gorge3* from the diploid progenitor. An additional peak not observed in the diploids was observed in the AD genome at a sequence diversity of 0.08 and was considered to be a result of a right shift of the extra 0.07 peak observed in the synthetic AD curve (Fig. 2A, synthetic AD versus the AD genome). One possible cause for this observation is that the *Gorge3* elements experienced further divergence since their reunification in the AD genome, which would ultimately lead to a right shift of the nucleotide diversity compared with the synthetic AD data. However, since no recent burst of *Gorge3* transposition is evident in the allopolyploid, this right shift may be due to a PCR bias that results in preferential amplification of younger elements in the diploid genomes, as indicated in Hawkins et al. (2009), or a more relaxed sampling of diverged sequences in the polyploid genome. Neither of these possibilities, however, impacts the key observation of no substantive *Gorge3* amplification since allopolyploid formation.

**Table 1.** Summary of PCR-amplified TE reverse transcriptase regions in *Gossypium.*

| | *gypsy*-like *Gorge3* | | *copia* | | LINE | |
|---|---|---|---|---|---|---|
| Species | No. of sequences | Nucleotide diversity[a] | No. of sequences | Nucleotide diversity[a] | No. of sequences | Nucleotide diversity[a] |
| *G. hirsutum* (AD genome) | 94 | 0.212 | 112 | 0.363 | 67 | 0.411 |
| *G. herbaceum* (A genome) | 235 | 0.183 | 143 | 0.057 | 66 | 0.394 |
| *G. raimondii* (D genome) | 114 | 0.302 | 148 | 0.271 | 65 | 0.381 |
| *G. exiguum* (K genome) | 211 | 0.193 | 129 | 0.107 | 101 | 0.391 |
| *Gossypioides kirkii* (outgroup) | 164 | 0.240 | 143 | 0.108 | 66 | 0.450 |
| Sum | 818 | | 675 | | 365 | |

[a]Pairwise divergence ($\pi$) among all sampled sequences.

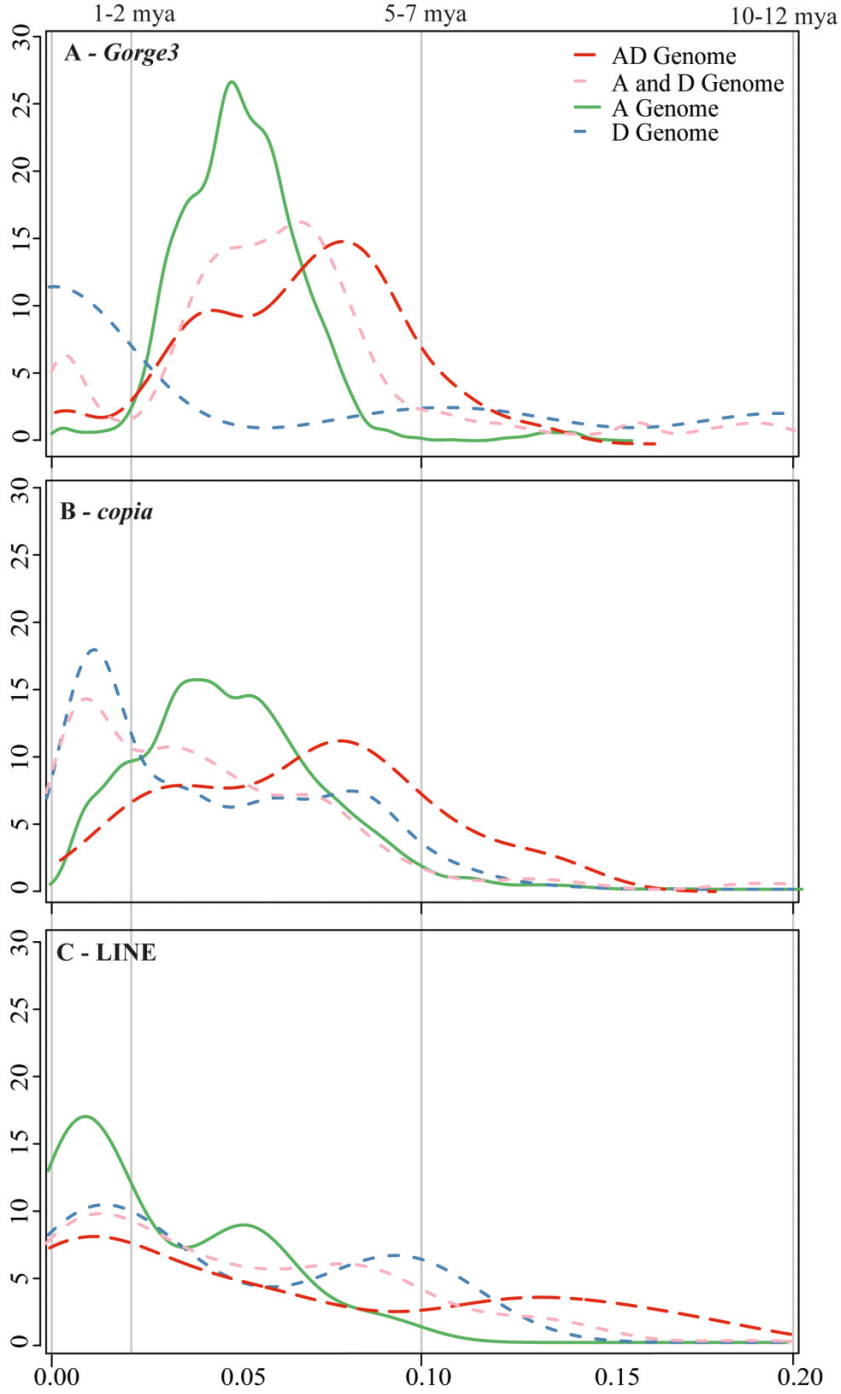
**Fig. 1.** Neighbor-joining analysis of PCR-amplified reverse transcriptase regions of transposable elements in *Gossypium* species and the phylogenetic outgroup *Gossypioides kirkii*. (A) *gypsy*-like *Gorge3* transposons, (B) *copia* retrotransposons, and (C) LINE retrotransposons. Taxa and genomes of origin are as follows: red, AD genome, *G. hirsutum*; green, A genome, *G. herbaceum*; purple, K genome, *G. exiguum*; blue, D genome, *G. raimondii*; and gold, *Gossypioides kirkii*. Sequences similar to transcripts found in the *G. hirsutum* EST database are indicated by gray shading (size is arbitrary).




### *copia* evolution in polyploid *Gossypium*

Degenerate primers from the RT regions of diverse *copia* elements were used to amplify a total of 112 sequences from the AD genome, each approximately 600 bp, which subsequently were aligned with 563 sequences from the diploid cotton genomes (A, 143; D, 148; K, 129) and 143 sequences from the outgroup *Gossypioides kirkii*. Similar to the analysis of *Gorge3* elements discussed above, phylogenetic analysis of *copia* sequences from the diploids and the AD genome revealed lineage-specific clades that could be used to assign a genome of origin to the sequences from allopolyploid cotton, resulting in 20 putative *copia* sequences of A-genome

**Fig. 2.** Timing of recent retrotransposon transpositions in *Gossypium*. The bottom axis represents the divergence for all pairwise comparisons among genome-specific sequences for each genome, the top axis is the estimated transposition time, and the *y*-axis is the density of pairwise comparisons at a given time point. Taxa and genomes are as follows: red, AD genome, *G. hirsutum*; pink, synthetic AD genome, combining sequences from the A and D genomes; green, A genome, *G. herbaceum*; and blue, D genome, *G. raimondii*.




origin and 6 of D-genome origin. Interestingly, a group of AD- and D-genome sequences, with slightly more than 10% average nucleotide divergence for the clade, clustered distantly from all the other *copia* RTs (Fig. 1B). With an average sequence diversity of 0.115, the radiation of this clade dates slightly earlier than the A- and D-genome speciation; however, the exclusion of all A-genome sequences suggests that this clade most likely represents D-genome-specific *copia* elements resulting from a burst of transposition shortly after the A–D divergence. If the 34 AD sequences in this clade are considered D-genome elements in origin, the *copia* composition of the AD genome appears biased toward elements that originated in the D genome.

Analysis of the *copia* transposition history in the AD genome suggests proliferation at about 2–3 and 4–5 mya, which corresponds to an increase in *copia* abundance ob-

served in the two parental diploid genomes (Fig. 2B). The increased density of the more ancient peak is likely due to the same factors that created a right shift in the *gypsy* data: high heterogeneity, mixed genome clades, and PCR bias.

### LINE evolution in polyploid *Gossypium*

Sixty-seven LINE RT sequences (380 bp) were amplified from the AD genome and aligned with 299 published LINE RT sequences from the diploid genomes (A, 66; D, 66; K, 101; and *Gossypioides kirkii*, 66). The phylogenetic tree derived from these sequences (Fig. 1C) revealed a highly heterogeneous collection of LINE retrotransposons, characterized by rare lineage-specific clades and with sequences from multiple species typically clustering together in small clades spreading across the tree. Of the 67 LINE sequences from the allopolyploid *G. hirsutum*, 18 were inferred to originate from the A genome and 11 from the D genome. In contrast to the LTR retrotransposons, recent proliferation of LINEs in both diploid and polyploid genomes at a modest level (Fig. 2C) is indicated by the presence of multiple small lineage-specific AD-genome clusters with very short branches.

### Transcriptional activity of TEs in cotton

We looked for possible expression of transposable elements by searching the *Gossypium* EST library for the presence of transcripts matching our PCR sequences through BLASTN. For the LTR retrotransposons, only 3 hits were recovered (GenBank accession Nos. DT464842.1, DN800971.1, DN801158.1) from *G. hirsutum* ESTs; these hits shared more than 80% sequence identity with 20 PCR-amplified *Gorge3* RT sequences (see supplementary data).[2] All three *Gorge3*-like ESTs were derived from EST libraries constructed from stress-treated plants (cycloheximide and drought stressed). Interestingly, all 20 potentially transcribed *Gorge3* RT sequences, when viewed on the phylogenetic tree (Fig. 1A), unexpectedly clustered in older "*Gossypium*-specific" clades rather than in "lineage-specific" clades that went through more recent transposition bursts. This suggests that these elements may have retained the ability to be transcribed, either autonomously or non-autonomously, despite being relatively old. No *G. hirsutum* ESTs were recovered for *copia* elements.

Compared with *Gorge3* and *copia*, more EST hits were recovered for non-LTR LINEs (GenBank acc. Nos. DW480155.1, DW483875.1, DT467619.1, DW483876.1, DT561901.1, DN804881.1, AI055502.1, DW241491.1, DW488622.1, DW488621.1, DT461584.1, AI054871.1, DW480156.1, DT461724.1) (see supplementary data). These LINE-like ESTs, unlike the ESTs recovered for *Gorge3*, were derived from libraries from both stress-treated and untreated plants. These findings suggest that non-LTR retrotransposons may be more actively transcribed than LTR retrotransposons in *G. hirsutum*.

## Discussion

The present study evaluates the diversity of *Gorge3*, *copia*, and LINE retrotransposons in the genome of allopolyploid cotton, which completes a characterization of these transposable elements in the genus *Gossypium* that was initiated by a study of three diploid species having varying genome sizes (Hawkins et al. 2008, 2009). As in the diploid genomes, the LTR and non-LTR retrotransposons from the polyploid genome show remarkably different levels of heterogeneity and histories of transposition. For the LTR retrotransposons, *Gorge3* and *copia*, the AD-genome sequences clustered together in large clades with their A- or D-genome antecedent elements in a genome-specific fashion. In contrast, the LINEs appear to be highly variable in all *Gossypium* genomes; instead of clustering in major lineage-specific clades, the sequences from the polyploid were distributed across the tree in small neighboring clades that contained sequences from either diploid genome.

The above observations are consistent with our expectation that the majority of TEs in allopolyploid cotton were inherited from the parental diploid genomes and remain closely related to their lineal orthologs. Consequently, we were able to assign the parental genome of origin to many TE sequences from the allopolyploid genome and thereby evaluate the composition of elements originating from different parental genomes. The compositional ratio and genome bias inferred for the *copia* and LINE retrotransposons (approximately 1:2 and 1:2 A:D, respectively) were consistent with previous estimates of their total copy numbers in the model diploid progenitor genomes (Hawkins et al. 2006); however, the compositional ratio for *Gorge3* in the polyploid, while biased in the expected direction, displayed far greater bias than suggested by data from the model diploid progenitors (36:1 A:D in this study versus 6:1 in Hawkins et al. 2006 and 9:1 in Hawkins et al. 2009). This increased bias could suggest a recent amplification of A-genome-origin *Gorge3* elements after allopolyploid formation; however, no significant transpositional bursts following polyploid formation were resolved from the temporal estimation of *Gorge3* activity (Fig. 2A).

Possible biological explanations for these conflicting data are that the excess A-genome lineage-specific elements accumulated in the actual (and extinct) ancestor to the A genome and then transmitted vertically into the AD genome upon polyploid formation, or that the D-genome lineage-specific elements were removed from the actual D-genome ancestor after divergence from the extant model species and prior to polyploid formation. A third possible explanation is that the D genome experienced a very recent burst of *Gorge3* transposition independent of polyploidy (D-genome peak at nucleotide divergence 0 in Fig. 2A). This could impact our observation of genome bias in 2 ways: (1) it increases the number of *Gorge3* elements in the diploid, thereby lowering the genome bias ratio, and (2) the primers used may be more effective on the diploid D genome, where there are many young elements to sample. A final technical possibility is that there is preferential amplification of A-genome elements in the allopolyploid because of a difference in starting template number between the two coresident genomes. More data are required to evaluate the various possible explanations for the biases.

Another facet of our study was to evaluate the potential activation or mobilization of transposable elements follow-

[2] Supplementary data for this article are available on the journal Web site (http://genome.nrc.ca).

ing allopolyploidization. It was previously reported that a *copia*-like element of A-genome origin colonized a chromosome of D-genome origin in the AD genome after polyploid formation (Hanson et al. 2000; Zhao et al. 1998). Despite such evidence of post-polyploidization TE activity, there was a lack of correlation between TE transposition and polyploidization in our data. By our dating methods, clades with sequence diversity < 0.02–0.03 represent possible TE proliferation that took place less than 1–2 mya (i.e., after polyploidization; Fig. 2); however, no significant activity of either *Gorge3* or *copia* was observed within this divergence range, which was unexpected but not surprising. Although the activities of TEs have long been implicated as coincident with genome hybridization and (or) polyploidization since McClintock (1984), observations similar to ours here have also been reported (Baumel et al. 2002; Kentner et al. 2003; Robinson et al. 2000). Notable among these are observations from synthetic tetraploid wheat, where the insertion rates of LTR retrotransposons appear to stay consistent compared with parental B and A genomes (Charles et al. 2008), and also from *Spartina*, where little evidence of TE transposition was observed for natural populations of both the recent allopolyploid *Spartina anglica* (Baumel et al. 2002) and two natural $F_1$ homoploid hybrids (Parisod et al. 2009), one of which ultimately gave rise to the allopolyploid. As indicated in the recent review by Parisod et al. (2010), the activation of TE transposition may be restricted to specific elements in certain allopolyploid species instead of being an inevitable consequence of allopolyploidization.

In addition to assessing the historical activities of retrotransposons, we searched the cotton EST database for transcripts of TE reverse transcriptase sequences. Because active retrotransposons depend on the expression of RT genes for replication via reverse transcription, the presence of RT ESTs would suggest potential for current or recent TE proliferation. In the genome of polyploid *Gossypium*, LTR retrotransposons account for about 92% of TEs, whereas LINEs comprise only 1.5% (Guo et al. 2008). In the EST database, in contrast, the LTR retrotransposons were much less common than LINEs: only 3 ESTs were recovered for *Gorge3* (from libraries from stressed plants) and none were recovered for *copia*, which is suggestive of the tight suppression of LTR retrotransposons. However, a variety of LINE-like sequences were recovered from EST libraries derived from plants from both stressed and normal growing conditions.

Since the inactivation of TEs is a function of epigenetic suppression, one explanation for the relatively enhanced expression of LINEs could be their preferential insertion in genic regions (Tikhonov et al. 1999). Indeed, recent evidence from the maize genome suggests that LINEs tend to be evenly distributed across chromosomes, whereas LTR retrotransposons tend to be clustered in close association with pericentromeric heterochromatic regions and often accumulate inside other LTR retrotransposons, especially in families with high copy numbers (Baucom et al. 2009; Schnable et al. 2009). Epigenetic evidence from *Arabidopsis thaliana* indicates that a trade-off exists between the suppression of TEs via methylation and the deleterious effects of reduced expression of neighboring genes (Hollister and Gaut 2009). As a result, TEs that tend to exist near genes are less likely to be methylated, whereas deleterious TEs existing near genes tend to be removed quickly. Evidence from rice indicates that areas of more frequent recombination (which are often genic regions) are weakly and positively correlated with non-LTR retrotransposon abundance (Tian et al. 2009). The biased expression of these sequence types could be a result of their genomic insertional preferences and perhaps their less costly mutagenic properties.

The genomic consequences of polyploidy are of considerable interest owing to the prevalence of the phenomenon. The TE fraction of the genome has the potential to be highly volatile with respect to both rapid expansion and rapid sequence loss, which are not mutually exclusive outcomes of hybridization and polyploidization. Evidence for proliferation of TEs upon hybridization and polyploidization is abundant (Kashkush et al. 2002, 2003; Liu and Wendel 2000; Petit et al. 2007, 2010; Shan et al. 2005; Ungerer et al. 2006, 2009); however, the converse (i.e., TE quiescence) has also been observed for a number of plants (Baumel et al. 2002; Charles et al. 2008; Kentner et al. 2003; Parisod et al. 2009, 2010; Robinson et al. 2000). Here we find no evidence for a quantitatively impressive proliferation of TEs following polyploidization in cotton, although previous evidence using FISH (Hanson et al. 2000) implicates at least a modest level of TE activity in allopolyploid cotton. Much remains to be learned about the methods of control imposed by genomes on transposition to understand and potentially predict the outcomes of merging two divergent genomes having differing TE complements and regulatory systems.

## Acknowledgements

We thank J. Grafenberg and R. Rapp for assistance with sequence and statistical analyses, respectively, and J. Udall for help with aspects of the work involving the cotton EST database. We gratefully acknowledge support from the National Science Foundation Plant Genome Program.

## References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**(17): 3389–-3402. PMID:9254694.

Baucom, R.S., Estill, J.C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.M., et al. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet. **5**(11): e1000732. doi:10.1371/journal.pgen.1000732. PMID:19936065.

Baumel, A., Ainouche, M., Kalendar, R., and Schulman, A.H. 2002. Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). Mol. Biol. Evol. **19**(8): 1218–1227. PMID:12140233.

Capy, P., Gasperi, G., Biémont, C., and Bazin, C. 2000. Stress and transposable elements: co-evolution or useful parasites? Heredity, **85**(2): 101–106. doi:10.1046/j.1365-2540.2000.00751.x. PMID:11012710.

Charles, M., Belcram, H., Just, J., Huneau, C., Viollet, A., Couloux, A., et al. 2008. Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. Genetics, **180**(2): 1071–1086. doi:10.1534/genetics.108.092304. PMID:18780739.

de Boer, J.G., Yazawa, R., Davidson, W.S., and Koop, B.F. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics, **8**(1): 422. doi:10.1186/1471-2164-8-422. PMID:18021408.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5): 1792–1797. doi:10.1093/nar/gkh340. PMID:15034147.

Grover, C.E., Kim, H., Wing, R.A., Paterson, A.H., and Wendel, J.F. 2007. Microcolinearity and genome evolution in the *AdhA* region of diploid and polyploid cotton (*Gossypium*). Plant J. **50**(6): 995–1006. doi:10.1111/j.1365-313X.2007.03102.x. PMID:17461788.

Grover, C.E., Hawkins, J.S., and Wendel, J.F. 2008. Phylogenetic insights into the pace and pattern of plant genome size evolution. Genome Dyn, **4**: 57–68. doi:10.1159/000126006. PMID: 18756077.

Guo, W., Cai, C., Wang, C., Zhao, L., Wang, L., and Zhang, T. 2008. A preliminary analysis of genome structure and composition in *Gossypium hirsutum.* BMC Genomics, **9**(1): 314. doi:10.1186/1471-2164-9-314. PMID:18590573.

Hanson, R.E., Islam-Faridi, M.N., Crane, C.F., Zwick, M.S., Czeschin, D.G., Wendel, J.F., et al. 2000. Ty1-*copia*-retrotransposon behavior in a polyploid cotton. Chromosome Res. **8**(1): 73–76. doi:10.1023/A:1009239522541. PMID:10730591.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., and Wendel, J.F. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium.* Genome Res. **16**(10): 1252–1261. doi:10.1101/gr.5282906. PMID:16954538.

Hawkins, J.S., Hu, G., Rapp, R.A., Grafenberg, J.L., and Wendel, J.F. 2008. Phylogenetic determination of the pace of transposable element proliferation in plants: *copia* and LINE-like elements in *Gossypium.* Genome, **51**(1): 11–18. doi:10.1139/G07-099. PMID:18356935.

Hawkins, J.S., Proulx, S.R., Rapp, R.A., and Wendel, J.F. 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc. Natl. Acad. Sci. U.S.A. **106**(42): 17811–17816. doi:10.1073/pnas.0904339106. PMID:19815511.

Holligan, D., Zhang, X., Jiang, N., Pritham, E.J., and Wessler, S.R. 2006. The transposable element landscape of the model legume *Lotus japonicus.* Genetics, **174**(4): 2215–2228. doi:10.1534/genetics.106.062752. PMID:17028332.

Hollister, J.D., and Gaut, B.S. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. **19**(8): 1419–1428. doi:10.1101/gr.091678.109. PMID: 19478138.

Hori, Y., Fujimoto, R., Sato, Y., and Nishio, T. 2007. A novel *wx* mutation caused by insertion of a retrotransposon-like sequence in a glutinous cultivar of rice (*Oryza sativa*). Theor. Appl. Genet. **115**(2): 217–224. doi:10.1007/s00122-007-0557-6. PMID:17492423.

Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., and Wessler, S.R. 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature, **431**(7008): 569–573. doi:10.1038/nature02953. PMID:15457261.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAS and their regulatory roles in plants. Annu. Rev. Plant Biol. **57**(1): 19–53. doi:10.1146/annurev.arplant.57.032905.105218. PMID:16669754.

Kashkush, K., Feldman, M., and Levy, A.A. 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics, **160**(4): 1651–1659. PMID:11973318.

Kashkush, K., Feldman, M., and Levy, A.A. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat. Genet. **33**(1): 102–106. doi:10.1038/ng1063. PMID:12483211.

Kentner, E.K., Arnold, M.L., and Wessler, S.R. 2003. Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana iris species and their use as molecular markers. Genetics, **164**(2): 685–697. PMID:12807789.

Labrador, M., Farré, M., Utzet, F., and Fontdevila, A. 1999. Interspecific hybridization increases transposition rates of *Osvaldo.* Mol. Biol. Evol. **16**(7): 931–937. PMID:10406110.

Lai, J., Li, Y., Messing, J., and Dooner, H.K. 2005. Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. Proc. Natl. Acad. Sci. U.S.A. **102**(25): 9068–9073. doi:10.1073/pnas.0502923102. PMID:15951422.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., et al.. 2001. Initial sequencing and analysis of the human genome. Nature, **409**(6822): 860–921. doi:10.1038/35057062. PMID:11237011.

Le, Q.H., Wright, S., Yu, Z., and Bureau, T. 2000. Transposon diversity in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. **97**(13): 7376–7381. doi:10.1073/pnas.97.13.7376. PMID: 10861007.

Liu, B., and Wendel, J.F. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome, **43**(5): 874–880. doi:10.1139/gen-43-5-874. PMID:11081978.

McClintock, B. 1984. The significance of responses of the genome to challenge. Science, **226**(4676): 792–801. PMID:15739260.

Morse, A.M., Peterson, D.G., Islam-Faridi, M.N., Smith, K.E., Magbanua, Z., Garcia, S.A., et al. 2009. Evolution of genome size and complexity in *Pinus.* PLoS One, **4**(2): e4332. doi:10.1371/journal.pone.0004332. PMID:19194510.

Neumann, P., Yan, H., and Jiang, J. 2007. The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. Genetics, **176**(2): 749–761. doi:10.1534/genetics.107.071902. PMID:17409063.

Noma, K., Ohtsubo, E., and Ohtsubo, H. 1999. Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. Mol. Gen. Genet. **261**(1): 71–79. doi:10.1007/s004380050943. PMID:10071212.

O'Neill, R.J., O'Neill, M.J., and Graves, J.A. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature, **393**(6680): 68–72. doi:10.1038/29985. PMID:9590690.

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics, **20**(2): 289–290. doi:10.1093/bioinformatics/btg412. PMID: 14734327.

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.A., and Ainouche, M. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina.* New Phytol. **184**(4): 1003–1015. doi:10.1111/j.1469-8137.2009.03029.x. PMID:19780987.

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., et al. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol. **186**(1): 37–45. doi:10.1111/j.1469-8137.2009.03096.x. PMID:20002321.

Petit, M., Lim, K.Y., Julio, E., Poncet, C., Dorlhac de Borne, F., Kovarik, A., et al. 2007. Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). Mol. Genet. Genomics, **278**(1): 1–15. doi:10.1007/s00438-007-0226-0. PMID:17375323.

Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q.T., et al. 2010. Mobilization of retrotransposons in synthetic allote-

traploid tobacco. New Phytol. **186**(1): 135–147. doi:10.1111/j.1469-8137.2009.03140.x. PMID:20074093.

Petrov, D.A., Schutzman, J.L., Hartl, D.L., and Lozovskaya, E.R. 1995. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis.* Proc. Natl. Acad. Sci. U.S.A. **92**(17): 8050–8054. doi:10.1073/pnas.92.17.8050. PMID: 7644536.

Pidoux, A.L., and Allshire, R.C. 2005. The role of heterochromatin in centromere function. Philos. Trans. R. Soc. Lond. B Biol. Sci. **360**(1455): 569–579. doi:10.1098/rstb.2004.1611. PMID: 15905142.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., et al. 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. **16**(10): 1262–1269. doi:10.1101/gr.5290206. PMID:16963705.

Robinson, T.J., Wittekindt, O., Pasantes, J.J., Modi, W.S., Schempp, W., and Morris-Rosendahl, D.J. 2000. Stable methylation patterns in interspecific antelope hybrids and the characterization and localization of a satellite fraction in the Alcelaphini and Hippotragini. Chromosome Res. **8**(7): 635–643. doi:10.1023/A:1009294226213. PMID:11117360.

SanMiguel, P., and Bennetzen, J.L. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. (Lond.), **82**(Suppl. 1): 37–44. doi:10.1006/anbo.1998.0746.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. 1998. The paleontology of intergene retrotransposons of maize. Nat. Genet. **20**(1): 43–45. doi:10.1038/1695. PMID:9731528.

Schmidt, T., Kubis, S., and Heslop-Harrison, J.S. 1995. Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta vulgaris* L.): LINEs and Ty1-*copia*-like elements as major components of the genome. Chromosome Res. **3**(6): 335–345. doi:10.1007/BF00710014. PMID:7551548.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science, **326**(5956): 1112–1115. doi:10.1126/science.1178534. PMID:19965430.

Schwarz-Sommer, Z., Leclercq, L., Göbel, E., and Saedler, H. 1987. Cin4, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J. **6**(13): 3873–3880. PMID:16453815.

Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., et al. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium.* Mol. Biol. Evol. **20**(4): 633–643. doi:10.1093/molbev/msg065. PMID:12679546.

Shan, X., Liu, Z., Dong, Z., Wang, Y., Chen, Y., Lin, X., et al. 2005. Mobilization of the active MITE transposons *mPing* and *Pong* in rice by introgression from wild rice (*Zizania latifolia* Griseb.). Mol. Biol. Evol. **22**(4): 976–990. doi:10.1093/molbev/msi082. PMID:15647520.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell, **136**(3): 461–472. doi:10.1016/j.cell.2008.12.038. PMID:19203581.

Swofford, D.L. 2001. PAUP*: Phylogenetic analysis using parsimony (and other methods). Version 4 [computer program]. Sinauer Associates, Sunderland, Mass.

Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J.L., Jackson, S.A., et al. 2009. Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? Genome Res. **19**(12): 2221–2230. doi:10.1101/gr.083899.108. PMID:19789376.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L., and Avramova, Z. 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. U.S.A. **96**(13): 7409–7414. doi:10.1073/pnas.96.13.7409. PMID:10377428.

Turcotte, K., Srinivasan, S., and Bureau, T. 2001. Survey of transposable elements from rice genomic sequences. Plant J. **25**(2): 169–179. doi:10.1046/j.1365-313x.2001.00945.x. PMID: 11169193.

Udall, J.A., Swanson, J.M., Haller, K., Rapp, R.A., Sparks, M.E., Hatfield, J., et al. 2006. A global assembly of cotton ESTs. Genome Res. **16**(3): 441–450. doi:10.1101/gr.4602906. PMID: 16478941.

Ungerer, M.C., Strakosh, S.C., and Zhen, Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr. Biol. **16**(20): R872–R873. doi:10.1016/j.cub.2006.09.020. PMID:17055967.

Ungerer, M.C., Strakosh, S.C., and Stimpson, K.M. 2009. Proliferation of Ty3/*gypsy*-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. BMC Biol. **7**(1): 40. doi:10.1186/1741-7007-7-40. PMID:19594956.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., et al. 2001. The sequence of the human genome. Science, **291**(5507): 1304–1351. doi:10.1126/science.1058040. PMID:11181995.

Wessler, S.R. 1996. Plant retrotransposons: turned on by stress. Curr. Biol. **6**(8): 959–961. doi:10.1016/S0960-9822(02)00638-3. PMID:8805314.

Wicker, T., and Keller, B. 2007. Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. Genome Res. **17**(7): 1072–1081. doi:10.1101/gr.6214107. PMID:17556529.

Wright, D.A., Ke, N., Smalle, J., Hauge, B.M., Goodman, H.M., and Voytas, D.F. 1996. Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana.* Genetics, **142**(2): 569–578. PMID:8852854.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and van der Knaap, E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science, **319**(5869): 1527–1530. doi:10.1126/science.1153040. PMID: 18339939.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., et al. 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc. Natl. Acad. Sci. U.S.A. **103**(51): 19581–19586. doi:10.1073/pnas.0607142103. PMID:17158798.

Zabala, G., and Vodkin, L.O. 2005. The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell, **17**(10): 2619–2632. doi:10.1105/tpc.105.033506. PMID:16141454.

Zhao, X.P., Si, Y., Hanson, R.E., Crane, C.F., Price, H.J., Stelly, D.M., et al. 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Res. **8**(5): 479–492. PMID:9582192.

Zofall, M., and Grewal, S.I. 2006. RNAi-mediated heterochromatin assembly in fission yeast. Cold Spring Harb. Symp. Quant. Biol. **71**(1): 487–496. doi:10.1101/sqb.2006.71.059. PMID:17381331.