Phylogenetic determination of the pace of transposable element proliferation in plants: *copia* and LINE-like elements in *Gossypium*

Jennifer S. Hawkins, Guanjing Hu, Ryan A. Rapp, Jessie L. Grafenberg, and Jonathan F. Wendel

Abstract: Transposable elements contribute significantly to plant genome evolution in myriad ways, ranging from local insertional mutations to global effects exerted on genome size through accumulation. Differential accumulation and deletion of transposable elements may profoundly affect genome size, even among members of the same genus. One example is that of *Gossypium* (cotton), where much of the 3-fold genome size variation is due to differential accumulation of one *gypsy*-like LTR retrotransposon, *Gorge3*. *Copia* and non-LTR LINE retrotransposons are also major components of the *Gossypium* genome, but unlike *Gorge3*, their extant copy numbers do not correlate with genome size. In the present study, we describe the nature and timing of transposition for *copia* and LINE retrotransposons in *Gossypium*. Our findings indicate that *copia* retrotransposons have been active in each lineage since divergence from a common ancestor, and that they have proliferated in a punctuated manner. However, the evolutionary history of LINEs contrasts markedly with that of the *copia* retrotransposons. Although LINEs have also been active in each lineage, they have accumulated in a stochastically regular manner, and phylogenetic analysis suggests that extant LINE populations in *Gossypium* are dominated by ancient insertions. Interestingly, the magnitude of transpositional bursts in each lineage corresponds directly with extant estimated copy number.

Key words: genome size, cotton, transposable elements.

Résumé : Les éléments transposables contribuent significativement à l'évolution des génomes chez les plantes de diverses façons allant de mutations insertionnelles locales jusqu'à des effets globaux exercés sur la taille des génomes via leur accumulation. L'accumulation et la délétion différentielles des éléments transposables peuvent affecter de façon marquée la taille du génome, même au sein d'un même genre. Le genre *Gossypium* (cotonnier) en est un bon exemple puisqu'une part importante de l'amplitude de la variation dans la taille des génomes (3×) est due à l'accumulation différentielle d'un rétrotransposon-LTR de type *gypsy*, *Gorge3*. Les rétrotransposons *copia* et LINE sans LTR constituent également une part importante du génome du genre *Gossypium*, mais contrairement à *Gorge3*, leur nombre de copies n'est pas corrélé à la taille du génome. Dans le présent travail, les auteurs décrivent la nature et le moment de la transposition des rétrotransposons *copia* et LINE chez le genre *Gossypium*. Les observations indiquent que les rétrotransposons *copia* ont été actifs au sein de chaque lignage depuis la divergence à partir d'un ancêtre commun et que leur nombre s'est accru de manière ponctuelle. Par contre, l'évolution des LINE diffère de façon marquée de celle des rétrotransposons *copia*. Bien que les LINE aient également été actifs au sein de chaque lignage, ils s'accumulent d'une manière stochastique régulière et une analyse phylogénétique suggère que les populations d'éléments LINE existants chez le genre *Gossypium* sont dominées par des insertions anciennes. Fait intéressant à noter, l'ampleur des explosions de transposons dans chaque lignage correspond directement avec le nombre estimé de copies actuelles.

Mots-clés : taille du génome, cotonnier, éléments transposables.

[Traduit par la Rédaction]

Introduction

Transposable elements (TEs) are extraordinarily diverse and prominent components of plant genomes. Originally thought to be "junk" or "selfish" DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980), TEs are now recognized to play an important role in genome evolution via disruptive insertions and TE-induced gene duplication (Brunner et al. 2005; Hoen et al. 2006; Jiang et al. 2004; Lai et al. 2005; Morgante et al. 2005; Wang et al. 2006; Zabala and Vodkin 2005; Zhang et al. 2005), effects on

Received 27 July 2007. Accepted 15 October 2007. Published on the NRC Research Press Web site at genome.nrc.ca on 13 December 2007.

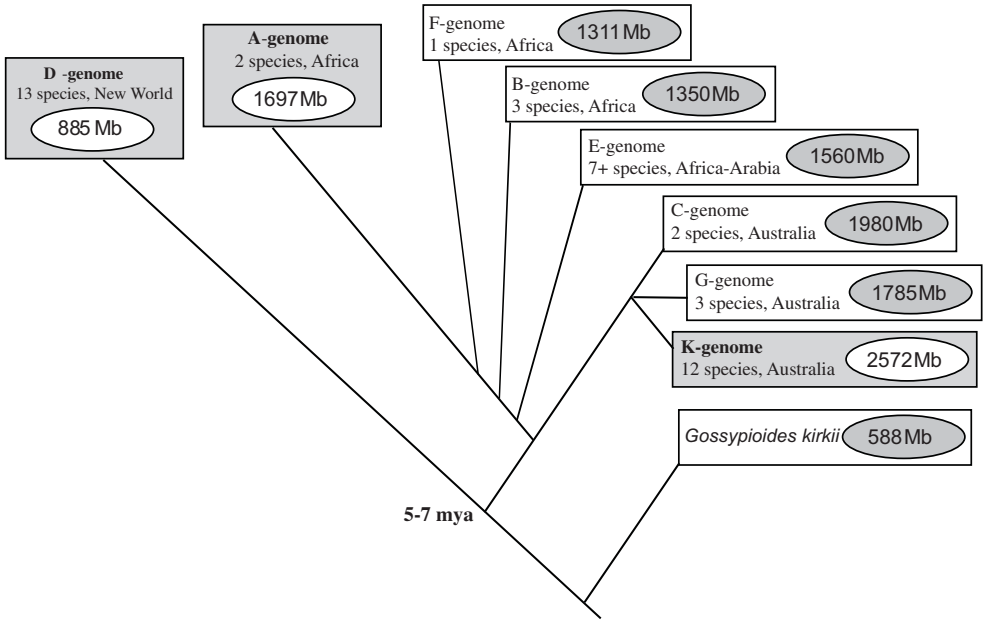
Corresponding Editor: T. Bureau.

J.S. Hawkins, G. Hu, R.A. Rapp, J.L. Grafenberg, and J.F. Wendel.[1] Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA.

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

 doi:10.1139/G07-099

Fig. 1. Genome sizes and evolutionary relationships among diploid members of *Gossypium* and a phylogenetic outgroup, *Gossypioides kirkii*. The indicated phylogenetic relationships are supported by several lines of molecular evidence (Seelanan et al. 1997; Small et al. 1998, 1999). *Gossypium* diverged from a common ancestor approximately 10–12 million years ago (mya) and experienced rapid radiation 5–7 mya (Cronn et al. 2002; Wendel and Cronn 2002). Genome sizes were reported by Hendrix and Stewart (2005).



gene expression (Hori et al. 2007; Lerat and Semon 2007; Yan et al. 2006), transposon-mediated chromosomal rearrangements (Gray 2000; Zhang and Peterson 1999), and their pronounced effects on genome size (Hawkins et al. 2006; Hill et al. 2005; Holligan et al. 2006; Neumann et al. 2006; Piegu et al. 2006; SanMiguel and Bennetzen 1998; SanMiguel et al. 1996). TEs, particularly LTR retrotransposons, may massively amplify and achieve extraordinarily high copy number within plant genomes over short periods of time. Over the past decade, it has become evident that most genome size variation in plants is due to the dynamic activity of LTR retrotransposons, in terms of differential rates of transposition (Hawkins et al. 2006; Piegu et al. 2006; SanMiguel and Bennetzen 1998) and variation in deletion rates through illegitimate recombination (Devos et al. 2002; Kirik et al. 2000; Ma et al. 2004) and intra-strand homologous recombination (Devos et al. 2002; Shirasu et al. 2000).

Consistent with the aforementioned studies, data from cotton (*Gossypium* spp.) indicate that LTR retrotransposons constitute a major fraction of all diploid species, regardless of total genome size (Hawkins et al. 2006). Analysis of genomic survey sequences from 3 *Gossypium* species whose genome sizes range approximately 3-fold suggests that most genome size variation among diploid members of the genus is due to differential accumulation of LTR retrotransposons and that different sequence types have proliferated in different genomes. Further analysis of the highly abundant *gypsy*-like retrotransposon family *Gorge3* showed that recent lineage-specific amplification has occurred in each of the species studied, although to various magnitudes and at different time points in each of their evolutionary histories (J.S. Hawkins et al., unpublished data). *Gorge3* proliferation was shown to occur in a punctuated manner in each of the genomes. However, as that study evaluated only the *Gorge3* family of LTR retrotransposons, it is presently unknown whether the results are applicable globally or whether they are specific to this particular sequence type.

To address this question, we investigate the evolutionary dynamics of the *copia* LTR retrotransposon and non-LTR LINE retrotransposons in *Gossypium*. Through PCR amplification of a portion of the reverse transcriptase domain from 3 representative diploid *Gossypium* species that range 3-fold in genome size and 1 phylogenetic outgroup, *Gossypioides kirkii* (Fig. 1), we evaluate the pace and tempo of transposition for each sequence type in each lineage. Results indicate that LTR retrotransposons accumulated in a punctuated manner at different times in different lineages. In contrast, LINEs appear to have experienced a low level of stochastically regular amplification with a few instances of elevated activity. We further show that the magnitude of transpositional bursts corresponds with previously estimated copy numbers for each of the sequence types.

Methods

Plant materials, DNA extraction, and phylogenetic analysis

Total genomic DNA was extracted from the A-genome species *G. herbaceum* (JMS; 1C = 1667 Mb), the D-genome species *G. raimondii* (JFW; 1C = 880 Mb), the K-genome species *G. exiguum* (Gos 5184; 1C = 2460 Mb), and the phylogenetic outgroup *Gossypioides kirkii* (JMS stock; 1C = 588 Mb), using young leaves and the DNeasy Plant Mini Kit (QIAGEN Inc., Valencia, California). Degenerate *copia* primers were designed by hand from aligned *Gossypium copia* genomic survey sequences (Hawkins et al. 2006) and subsequently tested to ensure amplification of the appropriate

sequence type with an acceptable range of sequence diversity. LINE primers were those described in Alix and Heslop-Harrison (2004). Primer sequences are as follows: *copia* Reverse, 5′-TGNTCCCAAATCTTTNATCTC-3′; *copia* Forward, 5′-GCNATGNANGANGAGATGGA-3′; LINE Forward, 5′-RVNRANTTYCGNCCNATHAG-3′; and LINE Reverse, 5′-GACARRGGRTCCCCCTGNCK-3′. The PCR reaction conditions were 94 °C for 3 min followed by 30 cycles of 94 °C for 1 min, 48 °C (*copia*) or 45 °C (LINE) for 2 min, and 72 °C for 1 min, ending with a final elongation step of 72 °C for 6 min. PCR products were purified, cloned using the pGEM®-T Easy Vector System (Promega, Wisconsin). Clones were sequenced using the T7 primer and BigDye® Terminator v3.2 Cycle Sequencing Kit (Applied Biosystems, California) at the Iowa State University DNA sequencing facility. Sequence data have been submitted to GenBank under accession Nos. EU102748–EU103609.

To confirm their identity, PCR sequences were queried using blastx to an *Arabidopsis copia* and LINE database provided by S. Wessler and F. Zhang (University of Georgia) and using tblastx to an in-house *Gossypium* database composed of previously identified genomic survey sequences. Confirmed sequences were aligned using MUSCLE (Edgar 2004) and were subsequently manually inspected. Neighbor-joining analyses were performed on the aligned sequences using PAUP* (Swofford 2003). Distances were set to the uncorrected (“p”) DNA/RNA distances and missing data were ignored for affected pair-wise comparisons.

Timing of lineage-specific transposition events

Based on evidence for orthologous A- and D-genome TEs in *Gossypium*, monophyletic TE clades sharing greater than 90% sequence identity are considered to be “lineage-specific” (Grover et al. 2007). Phrased alternatively, TEs with 90% or greater similarity are inferred to be related by duplication (transposition) events that occurred after divergence of the A-genome and D-genome lineages of plants (Fig. 1). Based on sequences from 48 nuclear genes, the genetic distance between *Gossypium* and the outgroup *Gossypioides kirkii* is approximately twice that of the A–D divergence (Senchina et al. 2003), which represents the basal-most split in the evolutionary history of *Gossypium* (estimated at 10 to 12 million years ago (mya); Cronn et al. 2002). Accordingly, TE lineages in *Gossypium* that shared an average of ≥80% sequence identity were considered to be *Gossypium*-specific. All other more divergent sequences were considered to have originated before the origin of the genus. To identify the lineage-specific PCR sequences, we constructed midpoint-rooted neighbor-joining trees based on pair-wise distances generated using Felsenstein’s 84 model. Pair-wise nucleotide diversity at each node was calculated, and the daughter sequences of all nodes with <0.1 average pair-wise divergence were extracted. Pair-wise nucleotide diversity (π) was calculated among all daughter sequences of each lineage-specific node, and density functions were plotted for the frequency of nucleotide divergences. Statistical analysis was performed in R (R Development Core Team 2005) with the base package, and with the phylogenetics package APE (Paradis et al. 2004).

Results

Copia evolution in *Gossypium*

PCR amplification and neighbor-joining analysis

A total of 563 (A = 143, D = 148, K = 129, *G. kirkii* = 143) *copia* reverse transcriptase (RT) sequences of approximately 600 bp in length were PCR amplified, aligned, and subjected to neighbor-joining analysis. The resulting phylogram (Fig. 2A) contained distinct, species-specific clusters of sequences for each genome in addition to several basal clades containing sequences from various, primarily cotton, genomes. There was high bootstrap support for the *G. kirkii*–specific clade (BS = 83) and the large *Gossypium*-specific clade (BS = 82). Additionally, one divergent group of D-genome *copia* sequences clustered distantly from the other RT sequences and contained 100% bootstrap support. Inspection of sequences belonging to this clade revealed a single synapomorphic amino acid deletion.

Pace and timing of transposition

To evaluate the lineage-specific transpositional nature and timing of *copia* retrotransposition in each genome, we employed a cotton-specific TE molecular clock estimated from orthologous TEs in *Gossypium* (Grover et al. 2007). We calculated π among all pair-wise comparisons of “lineage-specific” PCR sequences and translated the sequence identity between *copia* paralogs into an estimated transposition date (Fig. 3B). Only monophyletic lineages in which members shared an average of 90% identity or greater across unambiguously aligned sequence at the node representing the most recent common ancestor were considered lineage-specific clades. Similar to our previous findings for the *gypsy* retrotransposon family *Gorge3* (Fig. 3A), *copia* retrotransposons proliferate in a mostly episodic fashion, that is, in temporally compressed periods at various points in the history of the plant lineages (Fig. 3B). A burst of *copia* transposition occurred approximately 4 mya in *G. kirkii* and in the largest *Gossypium* genome (K), after which time there appears to have been little *copia* activity. In contrast, D-genome *copias* appear to have amplified recently (within the last 1 million years), but also rather continuously over the last 4–5 million years, as judged from the broad, flat peak extending from about 0.02 to 0.09 sequence diversity. Pair-wise comparisons among A-genome, lineage-specific clusters peak at approximately 2–3 mya, but as with the D genome, there appears to have been a lower level of *copia* activity since the origin of this lineage throughout most of its history. Interestingly, the transpositional timing for the D- and A-genome *copias* is similar to that found for *Gorge3* (Fig. 3A), although the magnitude of the transpositional bursts as measured by peak height (pair-wise density) varies among sequence types.

LINE evolution in *Gossypium*

PCR amplification and neighbor-joining analysis

Degenerate primers amplified a 380 bp region of the LINE reverse transcriptase domain. A total of 299 (A = 66, D = 66, K = 101, and *G. kirkii* = 66) LINE RT sequences

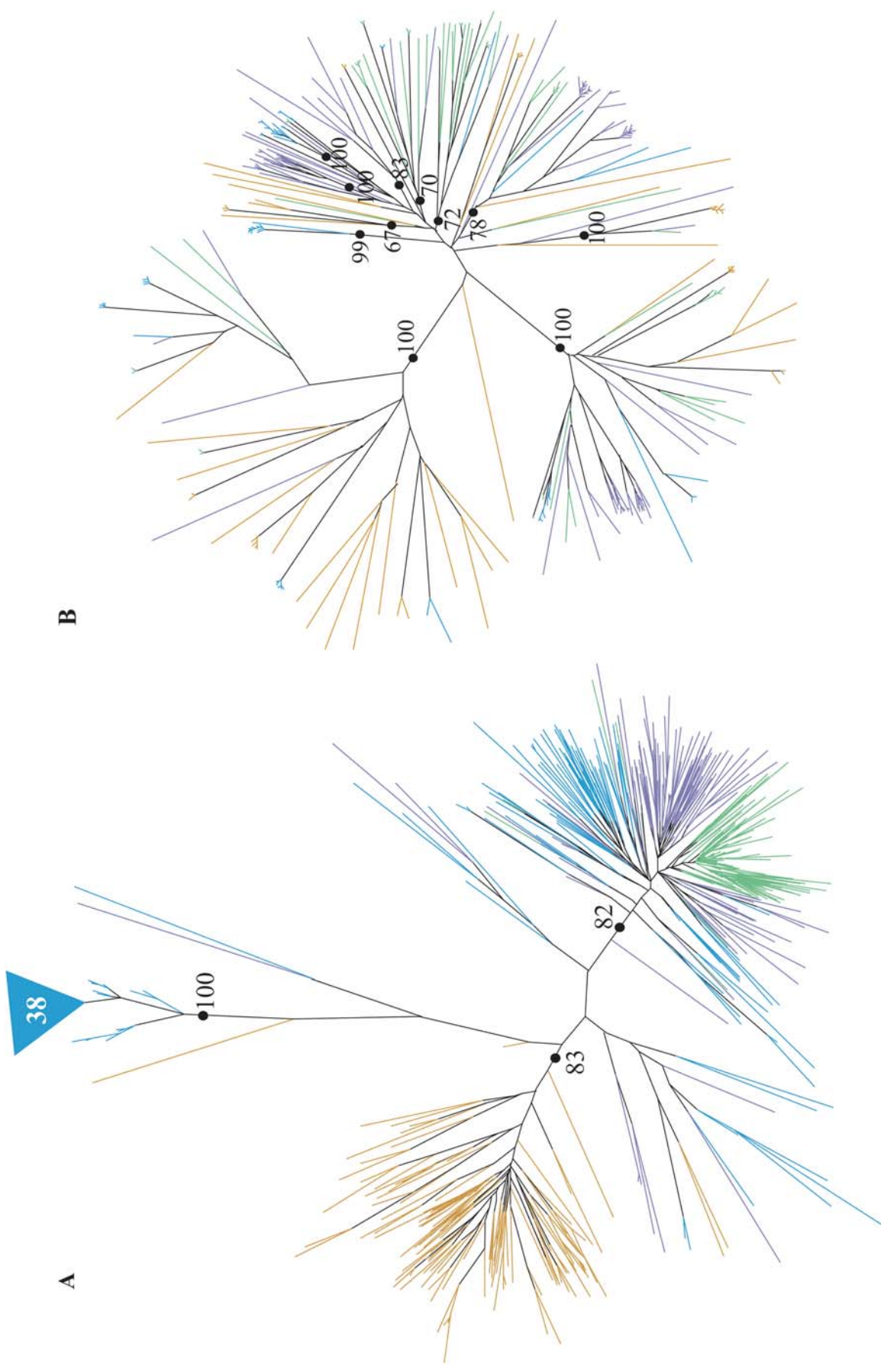
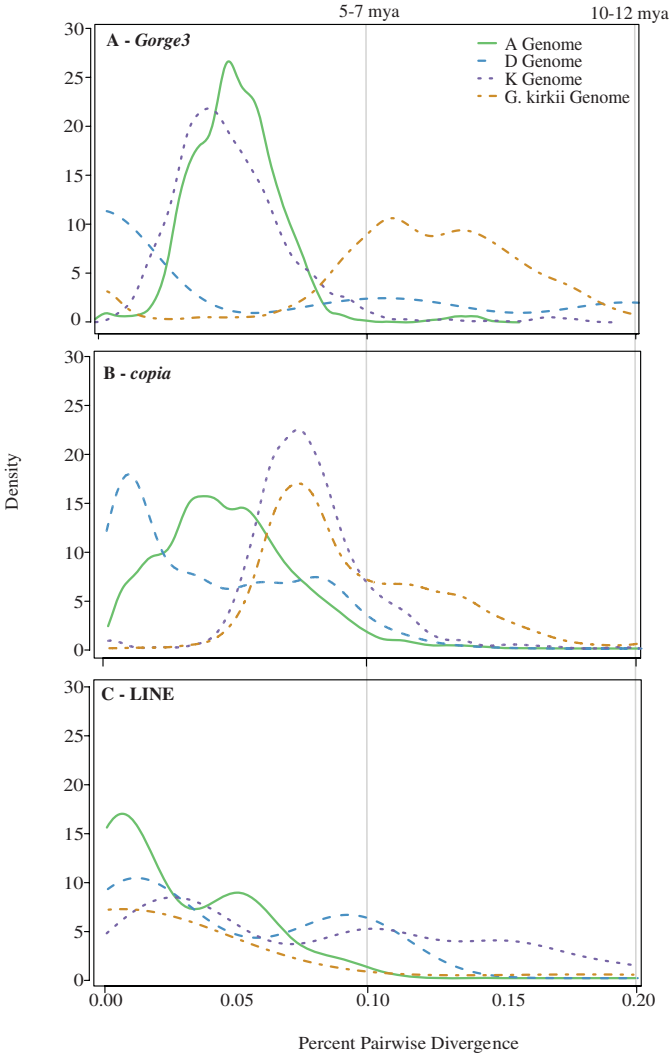


Fig. 2. Neighbor-joining analysis of PCR-amplified reverse transcriptase fragments of *copia* and LINE retrotransposons from *Gossypium* species and from a phylogenetic outgroup, *Gossypioides kirkii*. Panels A and B display the phylogenies for the *copia* and LINE neighbor-joining analyses, respectively. Bootstrap values for clades containing significant support are shown. The large blue triangle in panel A represents 38 recently diverged D-genome *copia* sequences. Orange, *Gossypioides kirkii*; blue, D genome, *G. raimondii*; green, A genome, *G. herbaceum*; and purple, K genome, *G. exiguum*.

Fig. 3. Timing of lineage-specific retrotransposon proliferation in *Gossypium*. Curves represent the distribution of pair-wise comparisons among lineage-specific sequences for each genome. The bottom axis shows the percentage sequence divergence between paralogs, the *y*-axis is the density of pair-wise comparisons, and the top axis is the estimated transposition time. (A) *Gypsy*-like *Gorge3*, (B) *copia*, (C) LINE.



from the A, D, K, and *G. kirkii* genomes were aligned and subjected to neighbor-joining analysis (Fig. 2B). In contrast with the *copia* and *Gorge3* phylogenies, LINE RTs clustered into what appears to be 3 major clades, each containing sequences from each of the 4 species studied. Two of the 3 clades contained bootstrap support of 100%, while the third large clade contained several internally supported groups. Branch lengths ranged from very long to small short terminal clusters, regardless of the genome from which they originated. Several small species-specific clusters for each genome are evident in all 3 major clades.

Pace and timing of transposition

The temporal profile of LINE transpositional events among lineage-specific sequences suggests, in contrast to that for LTR retrotransposons, stochastically regular transposition in all genomes, with a few punctuated periods of proliferation since divergence from a common ancestor (Fig. 3C). All genomes seem to have experienced recent transposition of LINEs and, unlike those of the LTR retrotransposons, the peak heights are mostly uniform among taxa, with the exception of some recent elevated activity in the A genome.

Discussion

Most of the genomic components responsible for the extraordinary genome size variation in plants are now clear, and the once apparently contradictory "C-value paradox" has graduated to the now perplexing "C-value enigma" (Gregory 2002, 2004). The significant impact of TE proliferation on genome size growth has been thoroughly documented, particularly with respect to large-scale sequencing projects in the grasses and a few model dicots. Recent, massive genomic bombardment by transposons has doubled the *Oryza australiensis* (Piegu et al. 2006) and maize (Meyers et al. 2001; SanMiguel and Bennetzen 1998; SanMiguel et al. 1996) genomes over only a few million years, and fewer than half of the TEs in the small genome of *O. sativa* originated before its origin, less than 680 000 years ago (Gao et al. 2004). Indeed, lineage-specific transposition has been observed in every plant genome investigated to date. Additionally, a recent investigation of *copia* evolution among Triticeae, rice, and *Arabidopsis* demonstrated that while large-genome taxa experience longer periods of transpositional activity, smaller genomes tend to purge these sequences at significantly faster rates (Wicker and Keller 2007). However, relatively little information exists with respect to the evolutionary dynamics of retrotransposon proliferation and removal among closely related genomes over short evolutionary time scales.

We show here that analysis of TE sequences within a phylogenetic framework yields novel temporal insights into patterns of transpositional activity for all major classes of retrotransposons. With this study and our analysis of the *gypsy* LTR retrotransposon family *Gorge3* (J.S. Hawkins et al., unpublished data), we have characterized the lineage-specific pace and timing of retrotransposon evolution among diploid members of *Gossypium*. These studies show that there has been recent, lineage-specific LTR retrotransposon activity, and that retroelements have proliferated in a punctuated fashion in all species studied, but in an idiosyncratic, lineage-specific fashion. Interestingly, bursts of *copia* and *Gorge3* amplification within the A and D genomes appear to have occurred at both similar and different time points during each species' evolutionary history. For example, the A genome experienced a burst of both *copia* and *Gorge3* activity approximately 2–3 mya, suggesting concurrent release from suppression for these two sequence types. However, although a burst of *Gorge3* and *copia* activity occurred in the D genome once within the last 1 million years, *copia* sequences persistently amplified over the previous 4 million years, while *Gorge3* was apparently inactive (Fig. 3B).

The pattern that emerges is one where there is transposition of different sequence types in each plant lineage, at different time points in each respective evolutionary history, and to varying levels of proliferation. The underlying causes behind punctuated proliferation are unclear, but it is com-

monly thought that bursts of transposition occur because of some form of biotic or abiotic stress, as organisms capable of genetic diversification under stress conditions are more likely to survive and reproduce (Wessler 1996). Hybridization has also been shown to induce TE amplification and proliferation (Liu and Wendel 2000; Ungerer et al. 2006) and, given that interspecific gene flow via hybridization is common in *Gossypium* (Cronn et al. 2003), may provide an explanation for the observed bursts of LTR retrotransposon activity.

The transpositional history of non-LTR LINEs contrasts remarkably with that of the LTR retrotransposons. LINEs appear to have experienced stochastically regular accumulation in each lineage, with a few peaks representing points of increased activity. No apparent periods of transpositional inactivity were evident. This suggests that LINEs propagate regularly over long evolutionary time periods in *Gossypium*, albeit at low levels, consistent with earlier results that showed relatively low copy numbers (Hawkins et al. 2006). In this respect, it is of interest to note that LINEs are more highly expressed in the *Gossypium* EST libraries than would be expected based on their estimated copy number (J.S. Hawkins, personal observation).

An interesting observation is the correspondence between the apparent magnitude of the transposition, as measured by peak height (density function), and estimated copy number from the whole-genome shotgun sequencing surveys (Hawkins et al. 2006). *Gorge3* copy numbers were found to be greatest in the two largest *Gossypium* genomes (A = 48181 ± 9257 and K = 88492 ± 12904), but very low and not significantly different from one another in the D genome (8674 ± 3683) and the outgroup *G. kirkii* (5502 ± 3305). Consistent with this observation, the magnitude of lineage-specific transposition is greatest in the A and K genomes, with little activity in the smaller genomes (Fig. 3A). Peak heights associated with the magnitudes of transpositional bursts of *copia* sequences exhibit a similar correspondence with previously estimated copy numbers (Fig. 3B): *copia* copy numbers were found to be highest in the largest and smallest *Gossypium* genomes (D = 57956 ± 9300 and K = 67700 ± 11324), slightly lower in the A genome (43181 ± 8774), and lowest in the outgroup (17006 ± 5765), while LINE copy numbers were greatest in the A (30000 ± 7335) and K (27563 ± 7271) genomes and slightly lower in the D (13011 ± 4503) and *G. kirkii* (16006 ± 5597) genomes. This correspondence was also reported for the rice species *Oryza australiensis* (Piegu et al. 2006).

Of particular relevance to the present work is the example of *gypsy* and LINE evolution in 3 diploid members of *Vicia* that vary in genome size (Hill et al. 2005). Combining their results with previous work describing *copia* retrotransposons in *Vicia*, the authors concluded that LTR retrotransposons experienced recent proliferation, as demonstrated by low levels of sequence diversity among PCR-amplified paralogs, but that LINEs were highly heterogeneous, indicating that LINE populations were dominated by ancient insertions. Additionally, *copia* copy number in *Vicia* is not correlated with genome size (Pearce et al. 1996), similar to our results in cotton. These similarities between *Gossypium* and *Vicia* with respect to retrotransposon evolution are remarkable, and suggest that there may be a degree of generality to our conclusions. The mirror images in two separate, phylogenetically distant plant groups suggest that TE life-history traits, such as intragenomic location and insertional target site preference, may play important roles with respect to the evolutionary pace of transposition and ability to accumulate.

Comparative studies such as this provide the underlying framework for understanding why some TEs are more successful than others and why some genomes are more permissive of TE proliferation than others. Given the rapid and recent nature of TE proliferation observed in various angiosperm systems, comparisons among closely related, recently diverged taxa are likely to provide the most precise information with respect to global patterns of TE evolution.

Acknowledgments

We thank R. Percifield for performing some data analysis, J. Swanson for help with computational analysis, and S. Wessler and F. Zhang for providing *copia* and LINE RT databases. This work was funded by the National Science Foundation.

References

Alix, K., and Heslop-Harrison, J.S. 2004. The diversity of retroelements in diploid and allotetraploid *Brassica* species. Plant Mol. Biol. **54**: 895–909. doi:10.1007/s11103-004-0391-z. PMID: 15612105.

Brunner, S., Pea, G., and Rafalski, A. 2005. Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize. Plant J. **43**: 799–810. doi:10.1111/j.1365-313X.2005.02497.x. PMID:16146520.

Cronn, R.C., Small, R.L., Haselkorn, T., and Wendel, J.F. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am. J. Bot. **89**: 707–725.

Cronn, R.C., Small, R.L., Haselkorn, T., and Wendel, J.F. 2003. Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. Evolution, **57**: 2475–2489. PMID: 14686525.

Devos, K.M., Brown, J., and Bennetzen, J.L. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis.* Genome Res. **12**: 1075–1079. doi:10.1101/gr.132102. PMID:12097344.

Doolittle, W., and Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature (London), **284**: 601–603. doi:10.1038/284601a0. PMID:6245369.

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

Gao, L., McCarthy, E.M., Ganko, E.W., and McDonald, J.F. 2004. Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. BMC Genomics, **5**: 18. doi:10.1186/1471-2164-5-18. PMID:15040813.

Gray, Y.H.M. 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet. **16**: 461–468. doi:10.1016/S0168-9525(00)02104-1. PMID:11050333.

Gregory, T.R. 2002. Genome size and developmental complexity. Genetica, **115**: 131–146. doi:10.1023/A:1016032400147. PMID: 12188045.

Gregory, T.R. 2004. Insertion-deletion biases and the evolution of genome size. Gene, **324**: 15–34. doi:10.1016/j.gene.2003.09.030. PMID:14693368.

Grover, C.E., Kim, H.R., 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**: 995–1006. doi:10.1111/j.1365-313X.2007.03102.x. PMID: 17461788.

Hawkins, J.S., Kim, H.R., 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**: 1252–1261. doi:10.1101/gr.5282906. PMID:16954538.

Hendrix, B., and Stewart, J.M. 2005. Estimation of the nuclear DNA content of *Gossypium* species. Ann. Bot. (Lond.), **95**: 789–797. doi:10.1093/aob/mci078. PMID:15701660.

Hill, P., Burford, D., Martin, D.M.A., and Flavell, A.J. 2005. Retrotransposon populations of *Vicia* species with varying genome size. Mol. Genet. Genomics, **273**: 371–381. doi:10.1007/s00438-005-1141-x. PMID:15891910.

Hoen, D., Park, K., Elrouby, N., Yu, Z., Mohabir, N., Cowan, R., and Bureau, T. 2006. Transposon-mediated expansion and diversification of a family of ULP-like genes. Mol. Biol. Evol. **23**: 1254–1268. PMID:16581939.

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**: 2215–2228. doi:10.1534/genetics.106.062752. PMID:17028332.

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**: 217–224. doi:10.1007/s00122-007-0557-6. PMID: 17492423.

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

Kirik, A., Salomon, S., and Puchta, H. 2000. 2000. Species-specific double-strand break repair and genome evolution in plants. EMBO J. **19**: 5562–5566. doi:10.1093/emboj/19.20.5562. PMID: 11032823.

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

Lerat, E., and Semon, M. 2007. Influence of the transposable element neighborhood on human gene expression in normal and tumor tissues. Gene, **396**: 303–311. doi:10.1016/j.gene.2007.04.002. PMID:17490832.

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

Ma, J., Devos, K.M., and Bennetzen, J.L. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. **14**: 860–869. doi:10.1101/gr.1466204. PMID:15078861.

Meyers, B.C., Tingey, S.V., and Morgante, M. 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res. **11**: 1660–1676. doi:10.1101/gr.188201. PMID:11591643.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., and Rafalski, A. 2005. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Genetics, **37**: 997–1002. PMID:16056225.

Neumann, P., Koblizkova, A., Navratilova, A., and Macas, J. 2006. Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics, **173**: 1047–1056. doi:10.1534/genetics.106.056259. PMID: 16585134.

Orgel, L., and Crick, F. 1980. Selfish DNA: the ultimate parasite. Nature (London), **284**: 604–607. doi:10.1038/284604a0. PMID: 7366731.

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

Pearce, S.R., Harrison, G., Dongtao, L., Heslop-Harrison, J.S., Kumar, A., and Flavell, A.J. 1996. The Ty1-*copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localization. Mol. Genet. Genomics, **250**: 305–315.

Piegu, B., Buyot, R., Picault, A., Saniyal, A., Kim, H.R., Collura, K., 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**: 1262–1269. doi:10.1101/gr.5290206. PMID:16963705.

R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

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**: 37–44. doi:10.1006/anbo.1998.0746.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science (Washington, D.C.), **274**: 765–768. doi:10.1126/science.274.5288.765. PMID: 8864112.

Seelanan, T., Schnabel, A., and Wendel, J.F. 1997. Congruence in the cotton tribe. Syst. Bot. **22**: 259–290. doi:10.2307/2419457.

Senchina, D., Alvarez, I., Cronn, R., Liu, B., Rong, J., Noyes, R., et al. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium.* Mol. Biol. Evol. **20**: 633–643. PMID: 12679546.

Shirasu, K., Schulman, A.H., Lahaye, T., and Schulze-Lefert, P. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res. **10**: 908–915. doi:10.1101/gr.10.7.908. PMID:10899140.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., and Wendel, J.F. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. **85**: 1301–1315. doi:10.2307/2446640.

Small, R.L., Ryburn, J.A., and Wendel, J.F. 1999. Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). Mol. Biol. Evol. **16**: 491–501. PMID: 10331275.

Swofford, D.L. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4 [computer program]. Sinauer Associates, Sunderland, Mass.

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

Wang, W., Zheng, H., Fan, C., Li, J., Shi, J., Cai, Z., et al. 2006. High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell, **18**: 1791–1802. doi:10.1105/tpc.106.041905. PMID:16829590.

Wendel, J.F., and Cronn, R.C. 2002. Polyploidy and the evolutionary history of cotton. Adv. Agron. **78**: 139–186.

Wessler, S.R. 1996. Plant retrotransposons: turned on by stress. Curr. Biol. **6**: 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**: 1071–1081.

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**: 19581–19586. doi:10.1073/pnas.0607142103. PMID:17158798.

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

Zhang, J., and Peterson, T. 1999. Genome rearrangements by nonlinear transposons in maize. Genetics, **153**: 1403–1410. PMID: 10545468.

Zhang, Y., Wu, Y., Liu, Y., and Han, B. 2005. Computational identification of 69 retroposons in *Arabidopsis*. Plant Physiol. **138**: 935–948. doi:10.1104/pp.105.060244. PMID:15923328.