

Evolutionary Dynamics of Transposable Elements Following a Shared Polyploidization Event in the Tribe Andropogoneae

Dhanushya Ramachandran,* Michael R. McKain,^{†*} Elizabeth A. Kellogg,[†] and Jennifer S. Hawkins*^{*,1}

*Department of Biology, 53 Campus Drive, West Virginia University, Morgantown, WV 26506, [†]Donald Danforth Plant Science Center, 975 North Warson Rd., St. Louis, MO 63132, and [‡]Department of Biological Sciences, 300 Hackberry Lane, The University of Alabama, Tuscaloosa, AL 35487

ORCID IDs: 0000-0002-9091-306X (M.R.M.); 0000-0003-1671-7447 (E.A.K.); 0000-0003-1216-5198 (J.S.H.)

ABSTRACT Both polyploidization and transposable element (TE) activity are known to be major drivers of plant genome evolution. Here, we utilize the *Zea-Tripsacum* clade to investigate TE activity and accumulation after a shared polyploidization event. Comparisons of TE evolutionary dynamics in various *Zea* and *Tripsacum* species, in addition to two closely related diploid species, *Urelytrum digitatum* and *Sorghum bicolor*, revealed variation in repeat content among all taxa included in the study. The repeat composition of *Urelytrum* is more similar to that of *Zea* and *Tripsacum* compared to *Sorghum*, despite the similarity in genome size with the latter. Although LTR-retrotransposons were abundant in all species, we observed an expansion of the *copia* superfamily, specifically in *Z. mays* and *T. dactyloides*, species that have adapted to more temperate environments. Additional analyses of the genomic distribution of these retroelements provided evidence of biased insertions near genes involved in various biological processes including plant development, defense, and macromolecule biosynthesis. Specifically, *copia* insertions in *Zea* and *T. dactyloides* were significantly enriched near genes involved in abiotic stress response, suggesting independent evolution post *Zea-Tripsacum* divergence. The lack of *copia* insertions near the orthologous genes in *S. bicolor* suggests that duplicate gene copies generated during polyploidization may offer novel neutral sites for TEs to insert, thereby providing an avenue for subfunctionalization via TE insertional mutagenesis.

KEYWORDS

adaptation
LTR
retrotransposon
copia insertions
maize
domestication
Tripsacum
dactyloides

Transposable element (TE) activation and accumulation generates significant genetic variation that can confer a range of effects on genome structure and function. As TEs carry ‘ready-to-use’ *cis*-elements, their insertions can impact gene regulation on a genome-wide scale by providing assorted regulatory elements to the adjacent genes. The new regulatory elements offered by inserted TEs can amplify and/or redistribute transcription factor binding

sites, thereby creating new regulatory networks or even participate in re-wiring of pre-existing networks (Hènaff *et al.* 2014, Lavalie *et al.* 2013, Krupovic *et al.* 2014, Huang *et al.* 2016, Carmona *et al.* 2016, Joly-Lopez *et al.* 2016). Several empirical studies have demonstrated TE-induced phenotypic changes associated with domestication and/or diversification of cultivated plants, including rice, maize, wheat, soybean, melon, and palm (Xiao *et al.* 2008, Naito *et al.* 2009, Fernandez *et al.* 2010, Studer *et al.* 2011, Uchiyama *et al.* 2013, Sanseverino *et al.* 2015, Ong-Abdullah *et al.* 2015, Lu *et al.* 2017). Indeed, TE-related polymorphisms are responsible for phenotypic variation in many agronomically important crops, demonstrating their importance in creating the genetic variability that contributes to plant genome evolution.

Hybridization, polyploidy, and stress have been shown to correlate with transposable element movement (Steward *et al.* 2000, Kalendar *et al.* 2000, Madlung *et al.* 2002, Ungerer *et al.* 2006, Ito *et al.* 2011, Cavrak *et al.* 2014, Bardil *et al.* 2015, Guo *et al.* 2017). Flowering plants are known to tolerate hybridization and polyploidy, both of which have promoted species diversification (Payseur and Rieseberg, 2016, Soltis *et al.* 2016, Goulet *et al.* 2017). These phenomena

Copyright © 2020 Ramachandran *et al.*

doi: <https://doi.org/10.1534/g3.120.401596>

Manuscript received March 8, 2020; accepted for publication September 21, 2020; published Early Online September 28, 2020.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.11954496>.

¹Corresponding author: Department of Biology, 53 Campus Drive, West Virginia University, Morgantown, WV 26506, E-mail: Jennifer.hawkins@mail.wvu.edu.

result in TE mobilization, leading to local mutations and genome size changes (Liu and Wendel 2000; Josefsson *et al.* 2006; Ungerer *et al.* 2006; Kawakami *et al.* 2010; Parisod *et al.* 2010; Piednoël *et al.* 2013). Furthermore, such bursts of TE activity result in insertional polymorphisms, often with deleterious effects on genome function; however, these effects could be nullified or shielded via gene duplication in polyploid genomes. Although the relationship between TE mobility and hybridization and/or polyploidization is unclear, it is speculated that such TE reactivation in response to genomic stresses could be due to incompatible suppression machinery between the two donor genomes, or that unknown mechanisms are in place that reduce genomic methylation under general stress conditions (Ha *et al.* 2009, Yaakov and Kashkush 2012, An *et al.* 2014, Senerchia *et al.* 2014, Defraia and Slotkin 2014, Ågren *et al.* 2016).

Previous studies of polyploidy in *Zea* have revealed evidence for a whole genome duplication (WGD) event at or shortly after the origin of grasses (Paterson *et al.* 2004), followed by another, more recent, WGD in the history of *Zea* that promoted the origin of the *Zea-Tripsacum* clade (Estep *et al.* 2014; McKain *et al.* 2018). Diversifying from a common ancestral allotetraploid ($n = 20$), both the *Zea* ($n = 10$) and *Tripsacum* ($n = 18$) genomes responded differentially to the diploidization process (Swignová *et al.* 2004, Schnable *et al.* 2009, Schnable & Freeling 2011). In addition to these chromosomal rearrangements, there is also evidence for retrotransposon invasion post divergence in both *Zea* and *Tripsacum* (Gaut *et al.* 2000). Hence, being divergent descendants of a common allopolyploid ancestor, the *Zea-Tripsacum* clade is a good model system to understand various evolutionary processes including the contribution of TEs to polyploidy, rediploidization, and species diversification.

Here, we describe TE activity and contribution to genome diversity in the *Zea-Tripsacum* clade that has undergone a recent shared polyploidization event (Figure 1A). We included two diploid relatives, *Urelytrum digitatum* and *Sorghum bicolor*, which provide an opportunity to explore TE-associated evolutionary events induced by

hybridization and genome doubling. By using clustering analysis, we have characterized the repetitive landscape in three *Zea* and three *Tripsacum* species (post allopolyploidization) and compared these results with that of the diploid relatives (pre-allopolyploidization). Our findings suggest recent and post-divergence activity of TEs in *Zea* and *Tripsacum* with lineage-specific expansions of *copia* elements specifically in *Z. mays* and *T. dactyloides*. Additionally, distributional analyses of newly expanded LTR-retrotransposons show insertional biases near genes involved in anatomical structure development, multicellular organismal processes, development, stress response, and defense.

MATERIALS AND METHODS

Plant material sources and Illumina sequencing of DNA

The following eight panicoid grasses were used in this study: *Zea mays* L., *Z. diploperennis* Iltis, Doebley & R. Guzmán, *Z. luxurians* (Durieu & Asch.) R.M. Bird, *Tripsacum dactyloides* (L.) L., *T. laxum* Nash, *T. australe* Cutler & E.S. Anderson, *Urelytrum digitatum* K.Schum., and *Sorghum bicolor* (L.) Moench. Short-read sequence data for *Zea mays* (SRS291653), *Zea luxurians* (SRR088692), *Tripsacum dactyloides* (SRS302460), and *Sorghum bicolor* (SRR5271056) were downloaded from the NCBI short read archive (Chia *et al.* 2012, Tenaillon *et al.* 2011, McCormick *et al.* 2018). Genomic short-read sequences of *Zea diploperennis*, *Tripsacum laxum*, *Tripsacum australe*, and *Urelytrum digitatum* were obtained from Dr. Elizabeth Kellogg, Donald Danforth Plant Science Center, St. Louis, Missouri and are deposited in (BioProject PRJNA664114).

Identification of TE families

Sequences were quality trimmed with Trimmomatic v0.33 (Bolger *et al.* 2014) using a sliding window of 4:25 and minimum length of 50 bp. Graph-based clustering of quality-trimmed reads was performed with RepeatExplorer, a pipeline designed to identify repeats

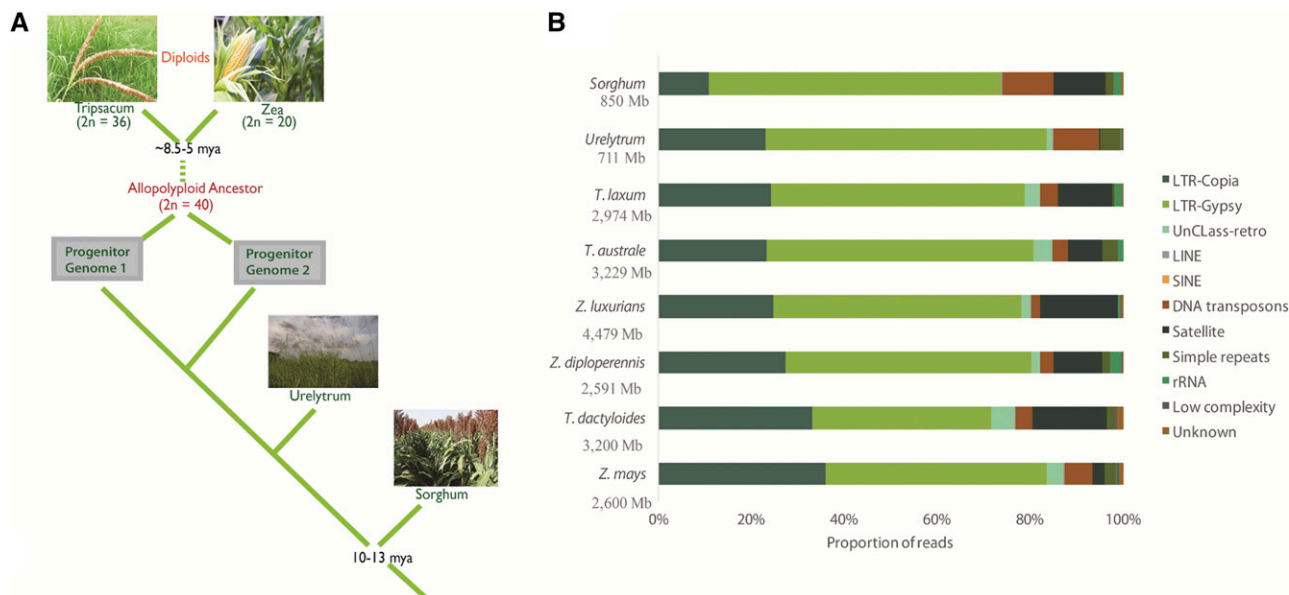


Figure 1 A. Evolutionary relationships of selected grass species, indicating polyplodization and species divergence. B. Proportional repeat composition. Genome size in Mb shown for each species in the y-axis. An expansion of *copia* families is observed in both *Z. mays* and *T. dactyloides* compared to related sister species. *Sorghum* displays a predominance of *gypsy* elements with a low level accumulation of *copia* families compared to *Zea-Tripsacum-Urelytrum* genomes.

from NGS reads (Novák *et al.* 2013). RepeatExplorer employs a clustering algorithm that quantifies similarities between all sequence reads and produces a graph that consists of nodes (sequence reads) and edges (connecting overlapping reads). Nodes are frequently connected to one another if they pass a threshold of 90% similarity over at least 55% of the sequence length, representing individual repetitive families.

Species-specific clustering analysis provides information regarding repeat quantities by reporting the number of reads per cluster, which can then be used to estimate the genome space occupied by each particular repeat, *i.e.*, (total length of each cluster (in Mb) x genome size (in Mb)) / total length of all clusters (in Mb) (Macas *et al.* 2015, Kelly *et al.* 2015, Ramachandran and Hawkins 2016). For species-specific clustering, three million reads (approximately 0.2x to 0.5x genome coverage) were sub-sampled from each dataset and processed to the format required by RepeatExplorer. Subsequently, all of the processed reads from all species were concatenated into one combined dataset, and the RepeatExplorer clustering was repeated in order to facilitate comparative analysis. All clusters were annotated using the Viridiplantae RepeatMasker library (<<http://www.repeatmasker.org>>) and categorized into repeat families. A plot representing interactions between repeat clusters among species was created using UpSetR (Lex *et al.* 2014).

Quantitative analysis of TE activity using molecular clock analysis

To estimate the timing of TE activity in each lineage, species-specific LTR sequences were extracted from each LTR-retroelement cluster. These species-specific reads were assembled using the Geneious *de novo* assembler to obtain a consensus sequence (Kearse *et al.* 2012). A grass-specific database was then used to extract LTRs from each consensus contig (blastn, e-value 1e-10, 85% identity). The best match for each species was chosen and the corresponding hit region was extracted using BEDTools v2.17.0 (Quinlan and Hall 2010).

To calculate LTR divergence (a rough measurement to estimate the age of a specific retrotransposon family), reads that were used for *de novo* assembly were mapped to the consensus LTR sequence using the Geneious reference genome assembler. To compare the LTRs of each species-specific repeat cluster, the percent identity was derived from the reference alignment by mapping each read to its respective LTR consensus sequence. Using a grass specific transposable element substitution rate of 1.3×10^{-8} per site per year (Ma and Bennetzen 2004), we estimated the activity of each major TE family in each species.

Genomic distribution of retroelements

Genome wide analyses for insertional bias of retroelements near protein coding genes in maize and sorghum were performed using the gene annotations for *Z. mays* (Zmays_284_5b+) and *S. bicolor* (Sbicolor_313_v3.1), downloaded from Phytozome v12.1 (Goodstein *et al.* 2012). The maize TE annotation file (B73v4.TE.filtered.gff3) was downloaded from Gramene Release 62 (<http://www.gramene.org>) and sorghum retroelement annotations were retrieved from our previous work described in Ramachandran and Hawkins 2016. BEDtools closest (version 2.17.0; Quinlan and Hall 2010) was used to identify LTR-retrotransposon (*gypsy* and *copia*) insertions within 1-5 kb upstream of genes. Insertion dating was performed as described above (Ma and Bennetzen 2004). Further, for genes associated with upstream retrotransposon insertions, gene set

enrichment analysis was performed in ShinyGO v0.6.1 (<http://bioinformatics.sdstate.edu/go/>; Ge *et al.* 2020).

As *Tripsacum* lacks a reference genome assembly, insertional bias analysis was performed using *Z. mays* reference gene annotations. *T. dactyloides* paired-end reads for which at least one read matched to our *gypsy* and *copia* clusters were identified using BLASTn v2.2.28+ (e-value = 1e-10, percent identity = 80). Matched reads were extracted and mapped to a library consisting of *copia* and *gypsy* contigs assembled by RepeatExplorer and to a filtered gene set containing the protein-coding genes from the *Z. mays* reference genome. Reference mapping of paired-end reads to the library was carried out using BWA aln v0.7.12 (Li and Durbin 2009) with the following optimized parameters: using the first 12 bases as a seed (-l 12), a maximum edit distance of four (-n 4), a maximum edit distance of two for the seed (-k 2), up to three gap openings (-o 3), a maximum of three gap extensions (-e 3), a mismatch penalty of two (-M 2), a gap opening penalty of six (-O 6), and a gap extension penalty of 3 (-E 3) (Mascagni *et al.* 2015). The results were used to generate a SAM file via the BWA "sampe" module, and then converted to a BAM file using SAMtools v1.9 (Li *et al.*, 2009). An LTR-retrotransposon was considered near a gene if one of the paired-end reads mapped to an LTR-retrotransposon and the other to a gene. The analysis was repeated for the remaining species included in the study (*Z. diploperennis*, *Z. luxurians*, *T. laxum*, *T. australe*, *U. digitatum*).

Phylogenetic analysis of retroelement families

To assess the evolutionary relationships of the shared *gypsy* and *copia* families, the reverse transcriptase (RT) and integrase (INT) amino acid domains were used for phylogenetic analysis. RepeatExplorer clusters were filtered for LTR-*gypsy* and *copia* elements with RT and INT domain blastx hits. RT reads were extracted from each cluster using the blastx output file and placed in separate genome-specific files. The reads were assembled for each cluster using the Geneious *de novo* assembler (Kearse *et al.* 2012). The resulting contigs were then confirmed to contain reverse transcriptase domains using blastx against the Cores-RT database (Llorens *et al.* 2011). RT sequences were then combined into a final query file for further analysis. The same analysis was performed for INT reads using Cores-INT database (Llorens *et al.* 2011).

Rpstblastn (e-value = 1e-10) was performed for the sequence dataset against the Conserved Domain Database (Marchler-Bauer *et al.* 2015) to identify and extract conserved regions. The best hits for each sequence were extracted, and the filtered blast output was converted to three-column bed format with matching coordinates for each hit. BEDTools v2.17.0 (Quinlan and Hall 2010) was used to extract the conserved regions (~540 bp for the RT domain and ~340 bp for the INT domain). The correct open reading frame from each sequence was identified using ORFfinder. All amino acid sequences were globally aligned with MUSCLE v3.8.31 (Edgar 2004). Alignments were manually inspected and adjusted in Bioedit v7.3.5 (Hall 1999). The optimal model of amino acid substitution for each alignment was estimated using Prot-test v3.4.5 according to the Akaike Information Criterion (AIC) (Abascal *et al.* 2005). In all cases except RT-*copia*, the best model selected was LG+G (Le and Gascuel. 2008). Blosum62+G was chosen as the optimal model for RT-*copia* (Henikoff and Henikoff. 1992). Likelihood analyses with 1,000 bootstrap replicates were performed in RAxML v.8.2.9 (Stamatakis *et al.* 2008) using the best model for each alignment. Bayesian analysis of alignments was performed in MrBayes v3.2.6 using rates = gamma and the respective substitution model (Ronquist and Huelsenbeck 2003). Two independent

■ **Table 1** Global repeat composition (in Mb) of species with respect to genome size. Genome size (in Mb) for each species is given below each species name. Estimated repeat content (in Mb) for each repeat family is listed below using individual repeat clustering analysis. Bold text represents the most abundant families in each genome.

	<i>Z. mays</i> (2600 Mb)	<i>Z. dip</i> (2591 Mb)	<i>Z. lux</i> (4479 Mb)	<i>T. dact</i> (3200 Mb)	<i>T. aus</i> (3229 Mb)	<i>T. lax</i> (2974 Mb)	<i>Urelytrum</i> (711 Mb)	<i>Sorghum</i> (848 Mb)
LTR/Copia								
<i>ji</i>	444.02	362.72	519.58	313.39	102.17	89.23	34.41	0.00
<i>opie</i>	327.96	224.42	534.88	444.17	286.28	356.65	0.00	0.00
<i>giepum</i>	45.85	26.29	22.95	84.73	20.09	48.38	7.13	13.74
<i>machiavelli</i>	12.07	3.73	3.95	3.17	0.00	6.61	0.00	0.00
<i>wiwa</i>	10.28	3.33	8.16	9.01	13.90	23.07	5.92	0.00
<i>uloh</i>	7.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>gudyeg</i>	7.03	0.00	2.79	0.00	0.00	0.00	0.00	0.00
<i>eninu</i>	2.44	2.35	0.00	5.15	6.86	9.71	0.00	0.00
<i>dijap</i>	2.27	2.63	0.00	145.58	240.43	159.12	0.00	0.00
<i>stonor</i>	1.22	0.00	0.00	6.06	2.71	4.62	0.00	0.00
<i>anar</i>	1.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>raider</i>	0.00	3.73	0.00	0.00	11.01	3.71	0.00	0.00
<i>fourf</i>	0.00	6.67	11.39	1.48	0.97	1.37	0.00	3.44
<i>ebel</i>	0.00	9.42	0.00	0.00	43.14	17.72	0.00	0.00
<i>maximus</i>	0.00	0.00	0.00	0.00	0.00	0.00	47.48	15.94
<i>gudyeg</i>	0.00	0.00	0.00	0.00	0.00	0.00	3.10	0.00
<i>gina</i>	0.00	0.00	0.00	0.00	0.00	0.00	2.79	0.00
<i>Angela</i>	0.00	0.00	0.00	0.00	0.00	0.00	1.58	1.66
<i>osr14</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.00
<i>Ale</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.48
<i>ruda</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00
<i>Os4</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.25
<i>nuhan</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.49
<i>ekasi</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.28
<i>Unclass.Copia</i>	70.06	67.44	4.20	35.85	13.57	62.07	63.53	40.61
Total	931.50	712.72	1107.90	1048.58	741.14	782.26	166.85	91.90
LTR/Gypsy								
<i>cinful-zeon</i>	289.37	223.63	583.02	58.70	80.87	65.06	0.00	0.00
<i>prem1</i>	220.81	70.62	308.55	170.72	51.08	104.28	0.00	0.00
<i>flip</i>	194.03	150.07	351.10	279.59	323.29	166.96	0.00	0.00
<i>gyma</i>	146.91	169.35	183.39	213.95	167.51	223.19	4.68	0.00
<i>xilon-diguus</i>	142.19	124.76	225.79	45.64	35.02	45.12	0.00	0.00
<i>tekay</i>	52.17	23.23	23.41	83.83	14.26	63.59	0.00	32.73
<i>uwum</i>	43.27	15.26	18.71	80.17	88.45	64.29	0.00	0.00
<i>grande</i>	36.39	80.04	136.62	0.00	86.64	44.80	0.00	0.00
<i>CRM4</i>	27.21	17.46	13.73	17.69	18.29	37.73	0.26	0.00
<i>dagaf</i>	20.93	47.00	50.30	30.26	55.60	119.78	0.00	0.00
<i>huck</i>	14.69	246.19	321.64	1.42	275.99	152.18	0.00	5.37
<i>doke</i>	7.55	74.35	78.71	0.00	39.71	46.03	0.00	0.00
<i>guhis</i>	4.53	4.12	9.87	1.41	0.94	1.59	0.00	9.71
<i>puck</i>	3.99	79.25	23.28	0.00	96.93	90.92	0.00	0.00
<i>lata</i>	0.00	0.00	0.00	17.19	0.00	44.18	0.00	0.00
<i>CRM1</i>	0.00	0.00	31.73	0.00	2.38	0.00	0.00	0.00
<i>kuni</i>	0.00	0.00	0.00	0.00	0.61	0.00	0.00	0.00
<i>Athila</i>	0.00	0.00	0.00	0.00	0.00	0.00	56.64	29.53
<i>Haight</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.59	0.00
<i>ahoru</i>	0.00	0.00	0.00	0.00	0.00	0.00	15.04	0.00
<i>leviathan</i>	0.00	0.00	0.00	0.00	0.00	0.00	18.95	66.39
<i>scDEL</i>	0.00	0.00	0.00	0.00	0.00	0.00	20.66	0.00
<i>scTat</i>	0.00	0.00	0.00	0.00	0.00	0.00	11.10	0.00
<i>m_457-190</i>	0.00	0.00	0.00	0.00	0.00	0.00	13.15	0.00
<i>evum</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	75.87
<i>retrosor6</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	181.60
<i>onap</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	40.13
<i>deiho</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	13.23
<i>retrosor1</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.17
<i>keama</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95

(continued)

■ **Table 1, continued**

	<i>Z. mays</i> (2600 Mb)	<i>Z. dip</i> (2591 Mb)	<i>Z. lux</i> (4479 Mb)	<i>T. dact</i> (3200 Mb)	<i>T. aus</i> (3229 Mb)	<i>T. lax</i> (2974 Mb)	<i>Urelytrum</i> (711 Mb)	<i>Sorghum</i> (848 Mb)
Unclass.Gypsy	37.05	93.91	29.59	226.53	488.30	492.92	296.70	77.75
Total	1241.09	1419.24	2389.45	1227.10	1825.87	1762.61	437.76	536.43
Unclass.	93.53	0.00	87.57	159.20	126.08	102.82	0.00	0.00
Retroelements								
LINE	2.72	0.00	1.00	2.91	0.00	0.00	1.22	0.94
SINE	0.00	0.00	0.00	0.00	0.00	1.60	0.15	0.00
DNA transposons								
DNA.CMC.EnSpm	80.22	61.68	77.51	85.00	88.81	107.27	36.06	48.78
DNA.MULE.MuDR	25.64	4.59	3.20	11.37	15.25	12.62	27.78	6.49
DNA.hAT.Ac	18.51	0.94	0.00	10.56	1.64	1.59	0.74	2.18
Helitron	17.78	9.44	6.75	2.28	0.00	0.00	0.35	1.78
DNA.PIF.Harbinger	14.96	0.00	1.29	4.51	0.00	3.62	2.47	29.57
DNA.TcMar.Stowaway	2.40	0.00	0.00	3.04	0.00	0.00	3.60	6.25
Total	159.90	76.64	88.75	116.75	105.70	125.09	71.00	95.05
Other Repeats								
Satellite	65.84	272.15	754.91	509.75	239.95	379.73	1.76	95.85
Simple repeats	63.14	46.10	0.00	52.94	107.11	13.80	31.42	12.58
rRNA	5.88	61.01	21.74	0.00	34.84	51.58	1.70	14.41
Low complexity	13.86	6.61	5.21	15.97	0.00	0.00	1.71	0.00
Unknown	20.72	3.65	20.38	43.23	0.00	7.41	1.19	4.03
Genome proportion	75%	74%	66%	66%	61%	60%	68%	61%

MCMC runs of 10 million generations were performed, sampling each run every 1,000 generations until convergence, with the potential scale reduction factor close to 1. All trees were visualized using FigTree v1.4.0.

Data availability

The authors affirm that all data necessary for confirming the conclusions of this article are represented fully within the article and its tables and figures. All sequence data are deposited as described in the Methods. Supplemental material available at figshare: <https://doi.org/10.25387/g3.11954496>.

RESULTS

Single-species clustering analysis

To evaluate repeat content with respect to genome size, we performed a separate clustering analysis for each species. Individual clustering allows the maximum number of reads to assemble, which increases the accuracy of the repeat estimates. We estimated the quantities of each repeat family in the genome using the following equation: (total length of each cluster (in Mb) x genome size (in Mb)) / total length of all clusters (in Mb) (Macas *et al.* 2015, Kelly *et al.* 2015). The estimated repeat compositions are shown in Table 1.

As expected, LTR-retrotransposons are the most abundant repeat in all species. Although the *Zea* species used in this study are diploid and contain the same number of chromosomes, the genome size of *Z. luxurians* (~4,479 Mb) is nearly double the size of the other two *Zea* species (~2,600 Mb). From the clustering analysis, *copia* elements were found to contribute approximately 710 Mb, 930 Mb, and 1,110 Mb to the *Z. diploperennis*, *Z. mays* and *Z. luxurians* genomes, respectively. *Gypsy* elements account for ~1,240 Mb and 1,420 Mb of the *Z. mays* and *Z. diploperennis* genomes, respectively, whereas ~2,390 Mb of the *Z. luxurians* genome is comprised of

gypsy elements (Table 1). The greater repeat abundance in *Z. luxurians* correlates with its larger genome size. The *Tripsacum* species have genomes of similar size (~3,100 Mb) and chromosome number ($2n = 36$), in which *T. laxum* possesses the smallest genome (2,974 Mb). *Copia* elements occupied ~740 and 780 Mb in *T. australe* and *T. laxum* genomes, in contrast to 1,050 Mb in the *T. dactyloides* genome. Approximately 1,760 Mb and 1,825 Mb is composed of *gypsy* elements in *T. laxum* and *T. australe*, respectively, whereas the *T. dactyloides* genome contains 1,230 Mb of *gypsy* elements. *Gypsy* elements contributed to more of the genome space (~53–59%) compared to *copia* (22–28%) in all *Zea* and *Tripsacum* species except for *Z. mays* and *T. dactyloides*, where both *gypsy* and *copia* were nearly equally abundant (Figure 1). *Urelytrum* and *Sorghum* contain ~167 Mb and 92 Mb of *copia*, and 438 Mb and 536 Mb of *gypsy* elements, respectively.

DNA transposons were found to contribute only 2–6% to the *Zea* and 2–3% to the *Tripsacum* genomes, in contrast to 10–11% in *Urelytrum* and *Sorghum* (Figure 1B). Other groups of repeat elements such as satellite repeats made up a significant fraction of the genome in several species. Approximately 755 Mb of the *Z. luxurians* and 510 Mb of the *T. dactyloides* genomes were occupied by satellite repeats. Although *Urelytrum* and *Sorghum* contain genomes of similar size, the former is composed of only 1.76 Mb of satellite DNA whereas the latter contained ~100 Mb of satellite DNA.

Repeat families and their contribution to genome size

From the individual repeat clustering analysis, we identified 24 *copia* and 30 *gypsy* families. Among the 24 *copia* families, *Ji* was the most abundant family in both *Z. mays* (444 Mb) and *Z. diploperennis* (363 Mb), whereas *Opie* was the most abundant in *Z. luxurians* (535 Mb) and in all of the *Tripsacum* species (Table 1). Disproportionately large increases are observed for *Ji* and *Opie* in *Z. mays* and *T. dactyloides*, in agreement with the overall increase of the *copia*

superfamily observed for these two species. Interestingly, *Dijap* was estimated at 146-240 Mb in *Tripsacum*, but contributed very little to the genome size of *Zea* (0-2 Mb), indicating lineage-specific accumulation of this family in the *Tripsacum* genus.

Among the *gypsy* families, *Cinful-Zeon*, *Prem1*, *Flip*, *Gyma*, *Huck* and *Xilon-Diguus* were abundant in both *Zea* and *Tripsacum*. The *Cinful-Zeon* family ranges from 224 - 583 Mb in *Zea* and is positively correlated with genome size, with the greatest abundance in *Z. luxurians*; however, this family contributes only ~70 Mb to the genomes of the *Tripsacum* species. This is also true for the *Xilon-Diguus* family, with estimates ranging from 125 - 226 Mb in *Zea* and ~42 Mb in *Tripsacum*, indicating amplification in the lineage leading to *Zea* after divergence from *Tripsacum*. The *Huck* family

is estimated at 246 Mb in *Z. diploperennis*, 321 Mb in *Z. luxurians*, 152 Mb in *T. laxum*, and 276 Mb in *T. australe*; however, *Huck* occupies only ~15 Mb of the *Z. mays* genome and ~1.4 Mb of the *T. dactyloides* genome. Similarly, elements such as *Doke* and *Puck* were noticeably more abundant in *Z. diploperennis*, *Z. luxurians*, *T. laxum*, and *T. australe*, suggesting independent lineage-specific loss of these families in these *Z. mays* and *T. dactyloides*.

There were 13 *gypsy* families that were specific to *Urelytrum* and/or *Sorghum*. *Athila* and *Leviathan* were relatively abundant (~19 - 66 Mb) and were identified in both *Urelytrum* and *Sorghum*. Apart from these two families, the remaining 11 *gypsy* families were predominantly present in *Sorghum*, but present in low copy number in *Urelytrum*, or absent altogether. For example, *Retrosor6* is estimated at ~180 Mb in

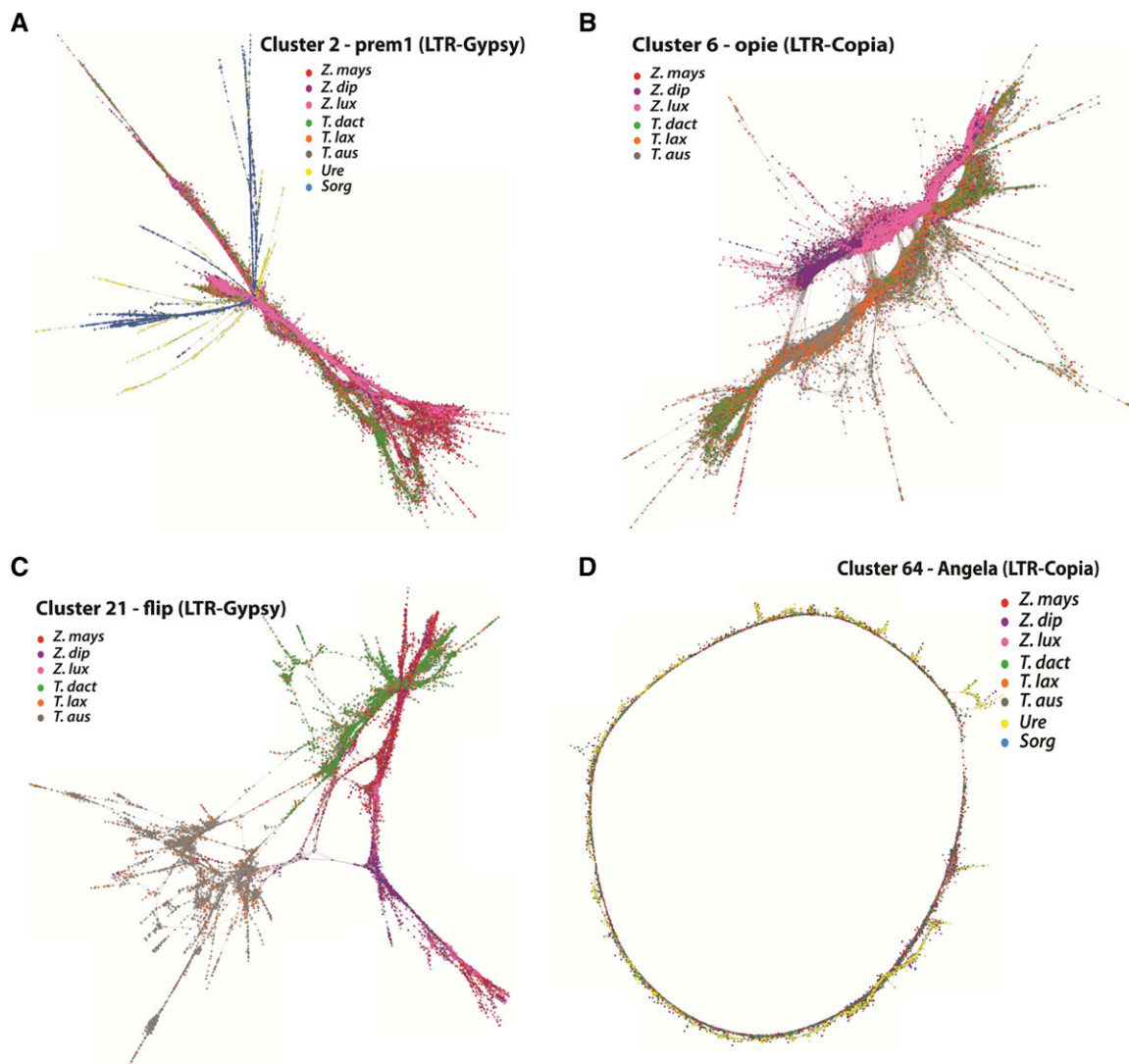


Figure 2 Comparative graph-based clustering. Graphs of individual repeat clusters that are shared between species demonstrating existing sequence variants within species. Highlighted dots represent sequences from individual species and lines connecting the dots represent sequence similarity. Each species is represented with a unique color: Red (*Z. mays*), purple (*Z. diploperennis*), pink (*Z. luxurians*), green (*T. dactyloides*), orange (*T. laxum*), and gray (*T. australe*). A. Cluster 2 shows shared LTR-gypsy elements in all genomes, in which sequences of *Zea* and *Tripsacum* are tightly connected with each other and sequences from *Urelytrum* and *Sorghum* are peripherally connected, concordant with their evolutionary relationships. B. Cluster 6 (*Opie*, LTR-copia) is an example of a lineage-specific repeat family, where sequences are shared between *Zea* and *Tripsacum*; however, there is a clear separation in clustering of both lineages. C. Cluster 21 (*Flip*, LTR-gypsy) shows three separate groups in which *Z. mays* and *T. dactyloides* are more similar to one another than either is to their sister species. D. Cluster 64 (*Angela*, LTR-copia) is an example of a tightly knitted graph in a linear arrangement shared between all eight species, demonstrating the conserved nature of ancient *Angela* elements across all included taxa.

Sorghum. Of the 48 *copia* clusters, only two were common to all species, ten were common to *Zea*, *Tripsacum* and *Urelytrum*, 19 clusters were exclusive to *Zea* and *Tripsacum*, and three were exclusive to *Urelytrum* and *Sorghum* (Figure S1B). Compared to *gypsy* super-families, there were fewer species-specific *copia* families.

Patterns of insertional bias

Investigation of the genomic distribution of retroelements in *Z. mays* uncovered 2,925 *gypsy* and *copia* elements inserted within 5 kb upstream of genes. The majority of these insertions were within 1-2 kb upstream of genes and were active within the last 1-3 million years (mya) (Figure S2 A and B). According to GO functional enrichment analysis, *gypsy* insertions were associated with genes involved in multicellular organismal process, development processes, and anatomical structure development (Figure S2 C), while *copia* insertions were associated with genes involved in response to abiotic stimulus and small molecule metabolic processes (Figure S2 D). Insertional biases in *Z. luxurians* and *Z. diploperennis* were similar to that of *Z. mays*, with enrichment of *copia* insertions near genes involved in abiotic stress response in addition to cellular, developmental, reproductive, and anatomical structure development (Figure S4, S5-A and B). Similarly, *copia* insertions in *T. dactyloides* were significantly enriched near genes involved in abiotic stress response in addition to processes involved in seed coat development. *Gypsy* insertions in *T. dactyloides* were enriched for starch, glucan and polysaccharide metabolic processes (Figure S3). In contrast, *copia* and *gypsy* insertions in *T. australe* were enriched near metabolic and biosynthetic processes, respectively, with no obvious bias toward genes involved in abiotic stress responses (Figure S4.C and S5.C). No significant enrichment for gene functional categories was found in either *T. laxum* and *U. digitatum*.

In *S. bicolor*, there were 511 *gypsy* and *copia* elements inserted within 5 kb upstream of genes. Most of these insertions were within 2-3 kb upstream of genes and were determined to be more recent insertional events (0-1 mya) compared to maize. In contrast to *Zea* and *Tripsacum*, both *gypsy* and *copia* insertions in *S. bicolor* were associated with genes involved in response to biotic stimulus, endoplasmic reticulum stress, and defense response to other organisms (Figure S6).

Evolutionary relationships and timing of transposition events

To assess the timing of major transposition events, we constructed maximum likelihood trees using the INT and RT domains (data not shown) of both *gypsy* and *copia* elements. Of the 127 shared *gypsy* clusters, 15 (total of 82 sequences) shared sufficient sequence identity within the integrase domain to allow amino acid sequence alignment. Major repeat families such as *Cinful-xeon*, *Prem1*, *Flip*, *Gyma*, *Xilon-Diguus*, and *Huck* were among these 15 clusters. With a few exceptions, most clades formed as expected in regard to species relationships. The *gypsy* families *Flip* and *Gyma* clustered together. The *Sorghum* and *Urelytrum* sequences from the *Flip* family clustered with *Zea* sequences of *Gyma*, whereas the *Zea* and *Tripsacum* sequences of *Flip* clustered with *Tripsacum* and *Sorghum* of *Gyma*. Several families such as *huck*, *puck*, and *grande* were clustered together with high support, suggesting a recent origin of these families. Clusters such as CL24 (unclassified), *uwum* (CL82) and *guhish* (CL132) also clustered with high sequence similarity.

We employed comparative sequence analyses of LTRs from 15 prominent clusters to estimate the temporal activity of retroelements

both pre- and post-divergence of the *Zea-Tripsacum* clade (Figure 4). The clusters chosen for this analysis are composed of the following repeat families: *Prem1*, *Flip*, *Cinful-Xeon*, *Gyma*, *Ji*, *Opie*, *Dijap*, *Retrosor-6*, and several prominent unclassified elements. In Figure 4A, the peak activity of each element per species per cluster is plotted against a TE-specific grass molecular clock (11 mya to present). The approximate timing of the *Zea-Tripsacum* divergence is highlighted in yellow (5-6 mya). *Zea* and *Tripsacum* have experienced post divergence lineage-specific activity for most repeat families. For example, *Ji*, *Opie*, and *Dijap* (CL7, CL12, CL15, CL42, CL19, and CL51) were active within the last three million years for all species in which they are present. The *Opie* element represented in CL7 is shared between *Zea*, *Tripsacum*, and *Urelytrum* and has been active within the last ~1-3 my indicating that amplification of *Opie* occurred in all three lineages (Figure 4A and 4B). In contrast, the amplification of CL2 (*prem1*) occurred recently only in *Z. luxurians* (2-3 mya) compared to all other species. Although *T. dactyloides* and *T. laxum* experienced increased activity of *Prem1* around the time of divergence, the activity of this element in *Z. mays*, *Z. diploperennis*, *T. australe*, and *Sorghum* dates as an older amplification event. Similarly, the activity of CL5 (*gyma*) is recent in *Z. diploperennis* but older in *Z. luxurians*. Several families were shared only between *Sorghum* and the wild relatives of *Zea* (*Z. diploperennis*, *Z. luxurians*) and *Tripsacum* (*T. laxum*, *T. australe*). Despite their presence, the activity of these families varies among species. For example, the activity of CL11 in *Z. luxurians* is recent (0-1 mya) but is dated as an older event in the other species (Figure 4A and 4C).

DISCUSSION

The present study evaluates TE dynamics in divergent descendants (*Zea* and *Tripsacum*) of a common allopolyploid ancestor within a phylogenetic framework and in comparison to two diploid relatives (*Urelytrum digitatum* and *Sorghum bicolor*). The comparative analysis of repeat elements from *Zea*, *Tripsacum*, *Urelytrum*, and *Sorghum* provides insight into the contribution of retrotransposons to genome evolution after a shared polyploidization event. Inclusion of additional *Zea* and *Tripsacum* species provided an opportunity to assess genomic variability in repeat content among species within a genus.

As expected, LTR-retrotransposons account for the majority of the repeats in the genomes of all species included in this study. Individual clustering analyses indicate that diverse LTR-retrotransposons contribute to genome size variation in this taxonomic group. Based on our comparative and molecular clock analyses, the majority of retrotransposon families are common to the *Zea - Tripsacum* clade in comparison to their diploid relatives, suggesting the proliferation of retrotransposons after allopolyploidization but before the split between the two genera (Figure 4A). Previous studies have hypothesized an occurrence of retroelement bursts just before the divergence of *Zea* and *Tripsacum* based on maize retroelement activity (Gaut *et al.* 2000, Estep *et al.* 2013). The results for the *Tripsacum* species included in the current analysis support this hypothesis, revealing a high number of shared retrotransposon families between the two genera. For example, *Ji* and *Opie* of the *copia* superfamily have been especially active (2 mya, Figure 4A, >300 Mb, Table 1) in both *Zea* and *Tripsacum*; however, these families contribute little (~35 Mb) to genome composition in *Urelytrum* and are absent in *Sorghum*. The presence and hyperactivity of these families in *Zea* and *Tripsacum* but not in *Sorghum* suggests amplification after the maize-sorghum split but before the allopolyploidization event leading to the *Zea-Tripsacum* lineage. Similarly, five *gypsy* families (*Cinful-Xeon*, *Prem1*, *Flip*,

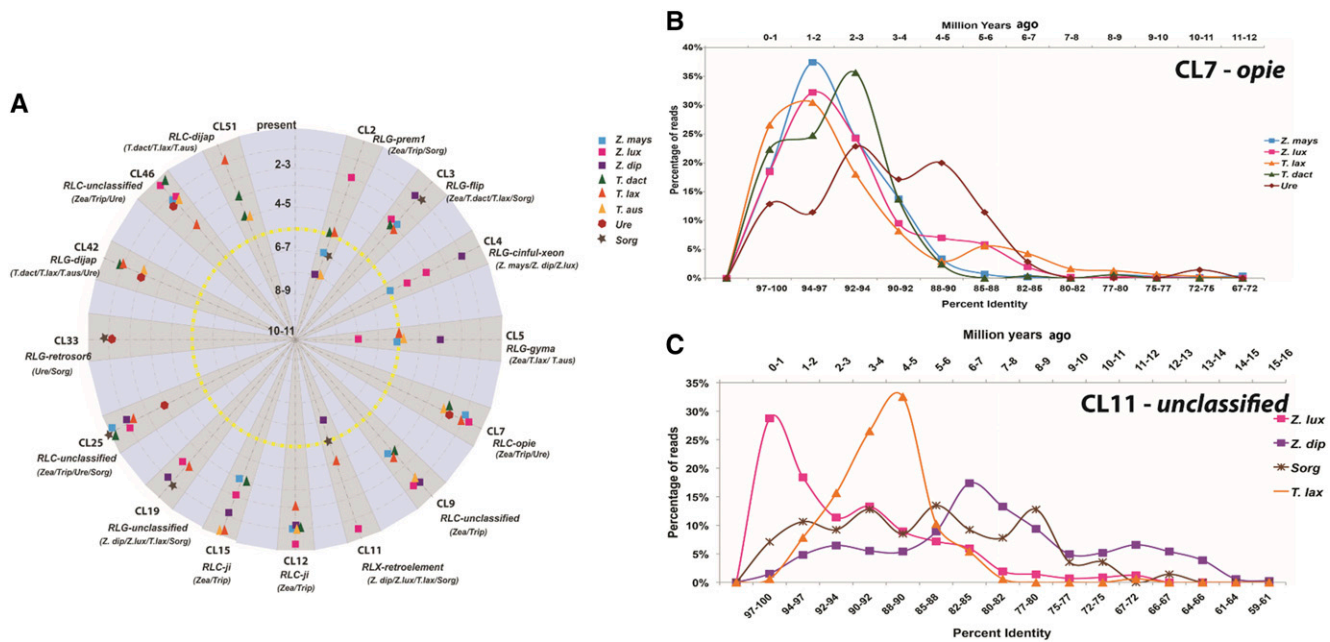


Figure 4 A. Activity of retroelements pre- and post-divergences of *Zea-Tripsacum* (yellow line) for 15 prominent retrotransposon clusters (shaded gray). Concentric circles indicate time scale per million years from 11 mya (center) to present (outer circle). For each cluster, the corresponding repeat family and shared species information is given below each cluster name. Each data point represents the peak activity of that element. B & C display retrotransposon activity of CL7 & CL11 based on percent identity of shared LTR sequences (bottom axis) and the corresponding grass molecular clock (mya) along the top axis. CL11 is absent in *Zea mays* and *Tripsacum dactyloides*.

Gyma, and *Huck*) are abundant in the *Zea-Tripsacum* clade but present in low copy numbers in the other lineages. Molecular clock analysis reveals recent activity (1-4 mya) for these families in *Zea* and *Tripsacum*, suggesting amplification after divergence from other taxa (Figure 4A). Conversely, families such as *Athila* and *Leviathan* have accumulated in the *Urelytrum* and *Sorghum* genomes, but are absent in *Zea* and *Tripsacum*, suggesting independent activation of LTR-retrotransposon families in different lineages over short evolutionary time scales.

Additionally, both *Zea* and *Tripsacum* contain lineage-specific families, indicating variation in retroelement amplification in each genus after divergence. Compared to *Zea*, a larger number of *gypsy* and *copia* clusters were unique to *Tripsacum* genomes. Seven *gypsy* families were common to all *Tripsacum* genomes and nine families were shared only between *T. laxum* and *T. australe* (Figure S1 & Figure 4A). The larger number of unique and recently active retroelements (~1-4 mya) in *Tripsacum* indicates the independent expansion of these families after the *Zea-Tripsacum* divergence. Overall, the abundance and recent activity of these genus-specific LTR-retrotransposons shared only between *Zea* and *Tripsacum* suggests that the activation of these families might be an outcome of shared polyploidization as proposed by the genomic shock hypotheses (McClintock 1984, Comai *et al.* 2003).

Surprisingly, clustering analyses suggest that *Z. mays* and *T. dactyloides* share greater similarity in TE composition and abundance for some retrotransposon families than either does to the other members of their respective genera. Examples of noticeably increased abundance in these two species include the families *Ji*, *Flip*, and *Opie*. Conversely, we also found significantly decreased copy numbers of *Huck*, *Doke*, and *Puck*. Also, a greater number of reads from both genomes are derived from the *copia* superfamily, indicating independent expansion of *copia* clades in both lineages. For a few shared

copia clusters, the peak of activity in *Z. mays* overlaps with *T. dactyloides*, suggesting both species experienced *copia* activity during a similar time period (Figure 4A). Considering the independent evolution of both species post divergence and the role of artificial selection in maize domestication, the similarity in composition and activity of *copia* elements in *Z. mays* and *T. dactyloides* suggests that natural and artificial selection have influenced TE amplification and accumulation similarly in both lineages. Although the precise cause of such species-specific activity is unclear, we propose that the observed patterns in TE abundance and accumulation may be related to adaptation to temperate climates. Indeed, recent evidence shows parallels in protein sequence evolution between natural and artificial selection in *T. dactyloides* and maize during their adaptation to temperate climates (Yan *et al.* 2019).

Considering the intensity of retroelement accumulation in maize as reported in other studies, it is likely that these elements have been active during maize domestication and improvement. Studies demonstrating TE involvement in plant domestication predominantly show that insertions near functional genes play a role in plant function and/or development. Well-known examples include *Hopscotch* involvement in apical dominance in maize (Studer *et al.*, 2011), *Gret1* in berry color variation in *Vitis vinifera* (Cadle-Davidson and Owens 2008), and LTR-mediated control of the blood orange phenotype (Butelli *et al.* 2012). Therefore, we explored the frequency of LTR-retrotransposon insertions near genes. In *Z. mays*, *Z. diploperennis*, and *Z. luxurians*, *gypsy* insertions were enriched near genes involved in multicellular, developmental, and metabolic processes whereas *copia* insertions were enriched near abiotic stress-associated and defense response genes (Figure S2-S5). In contrast with *Z. mays*, *copia* insertions in *Z. diploperennis* and *Z. luxurians* were also enriched near genes involved in developmental and metabolic processes. Indeed, *Z. diploperennis* and *Z. luxurians* shared similar GO enrichment profiles. *T. dactyloides* also

demonstrated enrichment of *copia* insertions near abiotic stress-associated genes, but *copias* were also enriched near genes involved in development, while *gypsy*s were enriched near genes involved in metabolic function, in contrast to findings in *Zea. T. australe* and *T. laxum* did not show enrichment for insertions near genes involved in the abiotic stress response. These results demonstrate that, with the exception of the insertional distributions in *Z. diploperennis* and *Z. luxurians*, TE insertional mutagenesis after polyploidization and divergence has been dynamic and lineage-specific.

The commonality of *copia* insertions near abiotic stress associated-genes in *Zea* and in *T. dactyloides* suggests that they may play a role in molding the plant response to environmental stress. Given our findings that *copias* have been particularly active in *Z. mays* and *T. dactyloides*, and that these insertions are predominantly near abiotic stress-associated genes, these results provide further support for a role during adaptation to temperate climates. Such insertional proximity provides the potential for TEs to affect the function of neighboring genes (Makarevitch *et al.* 2015, Cao *et al.* 2016, Pietzenek *et al.* 2016). For example, the tobacco Tnt1 and the rice Tos17 *copia* elements were found near stress-related genes, and the expression of these elements was linked with the biological responses of the plant to external stress (Grandbastien *et al.* 1997, Miyao *et al.* 2003, Le *et al.* 2007).

CONCLUSION

In this study, we provide insight into interspecific TE diversity and its contribution to genome evolution in related members of Andropogoneae that have undergone a shared polyploidization event. By including multiple accessions of two divergent species (*Zea* and *Tripsacum*) originating from a common allopolyploid ancestor, in addition to close diploid relatives (*Urelytrum digitatum* and *Sorghum bicolor*), we described LTR-retrotransposon diversity with respect to the hybridization and genome doubling process. Though the genome size of *Urelytrum* is similar to that of *Sorghum*, the repeat composition of *Urelytrum* is more like that of *Zea* and *Tripsacum*. Similarities in the proportion of the *copia* superfamily and satellite DNA in *Zea-Tripsacum-Urelytrum* suggests that *Urelytrum* or a close relative may have played an ancestral role in the origin of the *Zea-Tripsacum* lineages, supporting the hypothesis proposed by McKain *et al.* (2018). The genomic distribution of retroelements near protein coding genes of similar function in addition to the expansion of the *copia* superfamily exclusively in *Z. mays* and *T. dactyloides*, suggests amplification (and possibly participation) of new *copia* insertions during adaptation to temperate environments. Indeed, many of the *copia* insertions are near genes involved in abiotic stress and defense factors. Therefore, the *cis*-regulatory effects of TE insertions near genes may have influenced the evolution of *Z. mays* and *T. dactyloides* during both adaptation and domestication.

ACKNOWLEDGMENTS

We acknowledge the use of the Super Computing System Spruce Knob at West Virginia University, which is funded by the National Science Foundation EPSCoR Research Infrastructure Improvement Cooperative Agreement 1003907, the state of West Virginia (WVEPSCoR via the Higher Education Policy Commission) and West Virginia University.

LITERATURE CITED

Abascal, F., R. Zardoya, and D. Posada, 2005 ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105. <https://doi.org/10.1093/bioinformatics/bti263>

- Ågren, J. A., H.-R. Huang, and S. I. Wright, 2016 Transposable element evolution in the allotetraploid *Capsella bursa-pastoris*. *Am. J. Bot.* 103: 1197–1202. <https://doi.org/10.3732/ajb.1600103>
- An, Z., Z. Tang, B. Ma, A. S. Mason, Y. Guo *et al.*, 2014 Transposon variation by order during allopolyploidisation between *Brassica oleracea* and *Brassica rapa*. *Plant Biol (Stuttg)* 16: 825–835. <https://doi.org/10.1111/plb.12121>
- McKain, M. R., M. C. Estep, R. Pasquet, D. J. Layton, D. M. Vela Díaz *et al.*, 2018 Ancestry of the two subgenomes of maize. *bioRxiv*. <https://doi.org/10.1101/352351> (Preprint Posted June 20, 2018)<https://doi.org/10.1101/352351>
- Bardil, A., A. Tayalé, and C. Parisod, 2015 Evolutionary dynamics of retrotransposons following autopolyploidy in the Buckler Mustard species complex. *Plant J.* 82: 621–631. <https://doi.org/10.1111/tjp.12837>
- Bolger, A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Butelli, E., C. Licciardello, Y. Zhang, J. Liu, S. Mackay *et al.*, 2012 Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24: 1242–1255. <https://doi.org/10.1105/tpc.111.095232>
- Cadle-Davidson, M. M., and C. L. Owens, 2008 Genomic amplification of the Gret1 retroelement in white-fruited accessions of wild vitis and interspecific hybrids. *Theor. Appl. Genet.* 116: 1079–1094. <https://doi.org/10.1007/s00122-008-0737-z>
- Cao, C., J. Xu, G. Zheng, and X.-G. Zhu, 2016 Evidence for the role of transposons in the recruitment of cis-regulatory motifs during the evolution of C4 photosynthesis. *BMC Genomics* 17: 201. <https://doi.org/10.1186/s12864-016-2519-3>
- Carmona, L. M., S. D. Fugmann, and D. G. Schatz, 2016 Collaboration of RAG2 with RAG1-like proteins during the evolution of V(D)J recombination. *Genes Dev.* 30: 909–917. <https://doi.org/10.1101/gad.278432.116>
- Cavrak, V. V., N. Lettner, S. Jamge, A. Kosarewicz, L. M. Bayer *et al.*, 2014 How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet.* 10: e1004115. <https://doi.org/10.1371/journal.pgen.1004115>
- Chia, J.-M., C. Song, P. J. Bradbury, D. Costich, N. de Leon *et al.*, 2012 Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* 44: 803–807. <https://doi.org/10.1038/ng.2313>
- Comai, L., A. Madlung, C. Josefsson, and A. Tyagi, 2003 Do the different parental “heteromes” cause genomic shock in newly formed allopolyploids? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358: 1149–1155. <https://doi.org/10.1098/rstb.2003.1305>
- Defraia, C., and R. K. Slotkin, 2014 Analysis of retrotransposon activity in plants. *Methods Mol. Biol.* 1112: 195–210. https://doi.org/10.1007/978-1-62703-773-0_13
- Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Estep, M. C., J. D. DeBarry, and J. L. Bennetzen, 2013 The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. *Heredity* 110: 194–204. <https://doi.org/10.1038/hdy.2012.99>
- Fernandez, L., L. Torregrosa, V. Segura, A. Bouquet, and J. M. Martinez-Zapater, 2010 Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. *Plant J.* 61: 545–557. <https://doi.org/10.1111/j.1365-313X.2009.04090.x>
- Gaut, B. S., M. Le Thierry d'Ennequin, A. S. Peek, and M. C. Sawkins, 2000 Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. USA* 97: 7008–7015. <https://doi.org/10.1073/pnas.97.13.7008>
- Ge, S. X., Jung, D., and R. Yao, 2020 ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics.* 36: 2628–2659. <https://doi.org/10.1093/bioinformatics/btz931>
- Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes *et al.*, 2012 Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40: D1178–D1186. <https://doi.org/10.1093/nar/gkr944>

- Goulet, B. E., F. Roda, and R. Hopkins, 2017 Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiol.* 173: 65–78. <https://doi.org/10.1104/pp.16.01340>
- Grandbastien, M. A., H. Lucas, J. B. Morel, C. Mhiri, S. Vernhettes *et al.*, 1997 The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. *Genetica* 100: 241–252. <https://doi.org/10.1023/A:1018302216927>
- Guo, H., J. N. Mendrikahy, L. Xie, J. Deng, Z. Lu *et al.*, 2017 Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis. *Sci. Rep.* 7: 40139. <https://doi.org/10.1038/srep40139>
- Ha, M., J. Lu, L. Tian, V. Ramachandran, K. D. Kasschau *et al.*, 2009 Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proc. Natl. Acad. Sci. USA* 106: 17835–17840. <https://doi.org/10.1073/pnas.0907003106>
- Hall, T. A., 1999 Bioedit: A user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. https://doi.org/10.14601/Phytopathol_Mediterr-14998u1.29
- Hénaff, E., C. Vives, B. Desvoyes, A. Chaurasia, J. Payet *et al.*, 2014 Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species. *Plant J.* 77: 852–862. <https://doi.org/10.1111/tpj.12434>
- Henikoff, S., and J. G. Henikoff, 1992 Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 89: 10915–10919. <https://doi.org/10.1073/pnas.89.22.10915>
- Huang, S., X. Tao, S. Yuan, Y. Zhang, P. Li *et al.*, 2016 Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* 166: 102–114. <https://doi.org/10.1016/j.cell.2016.05.032>
- Ito, H., H. Gaubert, E. Bucher, M. Mirouze, I. Vaillant *et al.*, 2011 An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472: 115–119. <https://doi.org/10.1038/nature09861>
- Joly-Lopez, Z., D. R. Hoen, M. Blanchette, and T. E. Bureau, 2016 Phylogenetic and Genomic Analyses Resolve the Origin of Important Plant Genes Derived from Transposable Elements. *Mol. Biol. Evol.* 33: 1937–1956. <https://doi.org/10.1093/molbev/msw067>
- Josefsson, C., B. Dilkes, and L. Comai, 2006 Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16: 1322–1328. <https://doi.org/10.1016/j.cub.2006.05.045>
- Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo, and A. H. Schulman, 2000 Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* 97: 6603–6607. <https://doi.org/10.1073/pnas.110587497>
- Kawakami, T., S. C. Strakosh, Y. Zhen, and M. C. Ungerer, 2010 Different scales of Ty1/copia-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. *Heredity* 104: 341–350. <https://doi.org/10.1038/hdy.2009.182>
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung *et al.*, 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kelly, L. J., S. Renny-Byfield, J. Pellicer, J. Macas, P. Novák *et al.*, 2015 Analysis of the giant genomes of Fritillaria (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytol.* 208: 596–607. <https://doi.org/10.1111/nph.13471>
- Krupovic, M., K. S. Makarova, P. Forterre, D. Prangishvili, and E. V. Koonin, 2014 Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol.* 12: 36. <https://doi.org/10.1186/1741-7007-12-36>
- Lavialle, C., G. Cornelis, A. Dupressoir, C. Esnault, O. Heidmann *et al.*, 2013 Paleovirology of “syncytins”, retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368: 20120507. <https://doi.org/10.1098/rstb.2012.0507>
- Le, S. Q., and O. Gascuel, 2008 An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25: 1307–1320. <https://doi.org/10.1093/molbev/msn067>
- Le, Q. H., D. Melayah, E. Bonnivard, M. Petit, and M.-A. Grandbastien, 2007 Distribution dynamics of the Tnt1 retrotransposon in tobacco. *Mol. Genet. Genomics* 278: 639–651. <https://doi.org/10.1007/s00438-007-0281-6>
- Lex, A., N. Gehlenborg, H. Strobel, R. Vuillemot, and H. Pfister, 2014 UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* 20: 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liu, B., and J. F. Wendel, 2000 Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome* 43: 874–880. <https://doi.org/10.1139/g00-058>
- Llorens, C., R. Futami, L. Covelli, L. Domínguez-Escribá, J. M. Viu *et al.*, 2011 The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* 39: D70–D74. <https://doi.org/10.1093/nar/gkq1061>
- Lu, L., J. Chen, S. M. C. Robb, Y. Okumoto, J. E. Stajich *et al.*, 2017 Tracking the genome-wide outcomes of a transposable element burst over decades of amplification. *Proc. Natl. Acad. Sci. USA* 114: E10550–E10559. <https://doi.org/10.1073/pnas.1716459114>
- Ma, J., and J. L. Bennetzen, 2004 Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. USA* 101: 12404–12410. <https://doi.org/10.1073/pnas.0403715101>
- Macas, J., P. Novák, J. Pellicer, J. Čížková, A. Kobličková *et al.*, 2015 In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe Fabaeae. *PLoS One* 10: e0143424. <https://doi.org/10.1371/journal.pone.0143424>
- Madlung, A., R. W. Masuelli, B. Watson, S. H. Reynolds, J. Davison *et al.*, 2002 Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic Arabidopsis allotetraploids. *Plant Physiol.* 129: 733–746. <https://doi.org/10.1104/pp.003095>
- Makarevitch, I., A. J. Waters, P. T. West, M. Stitzer, C. N. Hirsch *et al.*, 2015 Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* 11: e1004915. <https://doi.org/10.1371/journal.pgen.1004915>
- Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz *et al.*, 2015 CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* 43: D222–D226. <https://doi.org/10.1093/nar/gku1221>
- Mascagni, F., E. Barghini, T. Giordani, L. H. Rieseberg, A. Cavallini, and L. Natali, 2015 Repetitive DNA and Plant Domestication: Variation in copy number and proximity to genes of LTR-retrotransposons among wild and cultivated sunflower (*Helianthus annuus*) genotypes. *Genome Biology and Evolution*. 7: 3368–3382. <https://doi.org/10.1093/gbe/evv230>
- McClintock, B., 1984 The significance of responses of the genome to challenge. *Science* 226: 792–801. <https://doi.org/10.1126/science.15739260>
- McCormick, R. F., S. K. Truong, A. Sreedasyam, J. Jenkins, S. Shu *et al.*, 2018 The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93: 338–354. <https://doi.org/10.1111/tpj.13781>
- Miyao, A., K. Tanaka, K. Murata, H. Sawaki, S. Takeda, *et al.*, 2003 Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *The Plant Cell.* 15: 1771–1780. <https://doi.org/10.1105/tpc.012559>
- Naito, K., F. Zhang, T. Tsukiyama, H. Saito, C. N. Hancock *et al.*, 2009 Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461: 1130–1134. <https://doi.org/10.1038/nature08479>
- Novák, P., P. Neumann, J. Pech, J. Steinhaisl, and J. Macas, 2013 RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29: 792–793. <https://doi.org/10.1093/bioinformatics/btt054>

- Ong-Abdullah, M., J. M. Ordway, N. Jiang, S.-E. Ooi, S.-Y. Kok *et al.*, 2015 Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* 525: 533–537. <https://doi.org/10.1038/nature15365>
- Parisod, C., K. Alix, J. Just, M. Petit, V. Sarilar *et al.*, 2010 Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* 186: 37–45. <https://doi.org/10.1111/j.1469-8137.2009.03096.x>
- Paterson, A. H., J. E. Bowers, and B. A. Chapman, 2004 Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. USA* 101: 9903–9908. <https://doi.org/10.1073/pnas.0307901101>
- Payseur, B. A., and L. H. Rieseberg, 2016 A genomic perspective on hybridization and speciation. *Mol. Ecol.* 25: 2337–2360. <https://doi.org/10.1111/mec.13557>
- Piednoël, M., G. Carrete-Vega, and S. S. Renner, 2013 Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. *Plant J.* 75: 699–709. <https://doi.org/10.1111/tpj.12233>
- Pietzenek, B., C. Markus, H. Gaubert, N. Bagwan, A. Merotto *et al.*, 2016 Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. *Genome Biol.* 17: 209. <https://doi.org/10.1186/s13059-016-1072-3>
- Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ramachandran, D., and J. S. Hawkins, 2016 Methods for accurate quantification of LTR-retrotransposon copy number using short-read sequence data: a case study in Sorghum. *Mol. Genet. Genomics* 291: 1871–1883. <https://doi.org/10.1007/s00438-016-1225-9>
- Ronquist, F., and J. P. Huelsenbeck, 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Sanseverino, W., E. Hénaff, C. Vives, S. Pinosio, W. Burgos-Paz *et al.*, 2015 Transposon Insertions, Structural Variations, and SNPs Contribute to the Evolution of the Melon Genome. *Mol. Biol. Evol.* 32: 2760–2774. <https://doi.org/10.1093/molbev/msv152>
- Schnable, J. C., and M. Freeling, 2011 Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS One* 6: e17855. <https://doi.org/10.1371/journal.pone.0017855>
- Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei *et al.*, 2009 The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115. <https://doi.org/10.1126/science.1178534>
- Senerchia, N., F. Felber, and C. Parisod, 2014 Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytol.* 202: 975–985. <https://doi.org/10.1111/nph.12731>
- Soltis, D. E., C. J. Visger, D. B. Marchant, and P. S. Soltis, 2016 Polyploidy: Pitfalls and paths to a paradigm. *Am. J. Bot.* 103: 1146–1166. <https://doi.org/10.3732/ajb.1500501>
- Stamatakis, A., P. Hoover, and J. Rougemont, 2008 A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57: 758–771. <https://doi.org/10.1080/10635150802429642>
- Steward, N., T. Kusano, and H. Sano, 2000 Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. *Nucleic Acids Res.* 28: 3250–3259. <https://doi.org/10.1093/nar/28.17.3250>
- Studer, A., Q. Zhao, J. Ross-Ibarra, and J. Doebley, 2011 Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* 43: 1160–1163. <https://doi.org/10.1038/ng.942>
- Swigonová, Z., J. Lai, J. Ma, W. Ramakrishna, V. Llaca *et al.*, 2004 Close split of sorghum and maize genome progenitors. *Genome Res.* 14: 1916–1923. <https://doi.org/10.1101/gr.2332504>
- Tenaillon, M. I., M. B. Hufford, B. S. Gaut, and J. Ross-Ibarra, 2011 Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biol. Evol.* 3: 219–229. <https://doi.org/10.1093/gbe/evr008>
- Uchiyama, T., S. Hiura, I. Ebinuma, M. Senda, T. Mikami *et al.*, 2013 A pair of transposons coordinately suppresses gene expression, independent of pathways mediated by siRNA in *Antirrhinum*. *New Phytol.* 197: 431–440. <https://doi.org/10.1111/nph.12041>
- Ungerer, M. C., S. C. Strakosh, and Y. Zhen, 2006 Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16: R872–R873. <https://doi.org/10.1016/j.cub.2006.09.020>
- Xiao, H., N. Jiang, E. Schaffner, E. J. Stockinger, and E. van der Knaap, 2008 A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319: 1527–1530. <https://doi.org/10.1126/science.1153040>
- Yaakov, B., and K. Kashkush, 2012 Mobilization of Stowaway-like MITEs in newly formed allohexaploid wheat species. *Plant Mol. Biol.* 80: 419–427. <https://doi.org/10.1007/s11103-012-9957-3>
- Yan, L., S. K. Kenchanmane Raju, X. Lai, Y. Zhang, X. Dai *et al.*, 2019 Parallels between natural selection in the cold-adapted crop-wild relative *Tripsacum dactyloides* and artificial selection in temperate adapted maize. *Plant J.* 99: 965–977.

Communicating editor: J. Birchler