

ORIGINAL ARTICLE

Different scales of *Ty1/copia*-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species

T Kawakami, SC Strakosh, Y Zhen and MC Ungerer
Division of Biology, Kansas State University, Manhattan, KS, USA

Activation of transposable elements in species' genomes represents an important mechanism of new mutation and of potential rapid change in genome size. Thus, it is increasingly recognized that transposable elements likely have played a significant role in shaping species' evolution. In an earlier report, we showed that the genomes of three sunflower species of ancient hybrid origin have experienced large-scale proliferation events of sequences within the *Ty3/gypsy*-like superfamily of long terminal repeat (LTR) retrotransposons. In this report, we investigate whether another superfamily of LTR retrotransposon (*Ty1/copia*-like elements) have experienced similar derepression and proliferation events in the genomes of these sunflower hybrid taxa. We show that *Ty1/copia*-like elements also have undergone

copy number increases following or associated with the origins of these species, although the scale of proliferation is less than that for *Ty3/gypsy*-like elements. Surveys of sequence heterogeneity of *Ty1/copia*-like elements in the genomes of the three hybrid and two parental species' genomes reveal that a single sub-lineage of these elements exhibits characteristics of recent amplification, and likely served as the proliferative source lineage. These findings indicate that the genomic and/or environmental conditions associated with the origins of these sunflower hybrid taxa were conducive to derepression of at least two major groups of transposable elements.

Heredity (2010) **104**, 341–350; doi:10.1038/hdy.2009.182; published online 13 January 2010

Keywords: transposable elements; genome evolution; *Ty1/copia*; hybridization

Introduction

Transposable elements are DNA sequences capable of mobilizing within the genome of their host (Bowen and Jordan, 2002; Feschotte *et al.*, 2002). In eukaryotes, these elements exist as a diverse array of DNA sequences capable of mobilization via either a 'copy-and-paste' (the class 1 elements) or 'cut-and-paste' (the class 2 elements) mechanism (Wicker *et al.*, 2007). These mechanisms represent fundamentally different modes of mobilization, use different molecular processes, and have different implications from the perspective of evolutionary changes in genome size. Recent years have witnessed renewed interest in the potential ability of transposable elements to impact the evolution of genes and genomes in plant and animal systems (McDonald, 1995; San-Miguel *et al.*, 1996; Kidwell and Lisch, 1997; Miller *et al.*, 1999; Bennetzen, 2000; van de Lagemaat *et al.*, 2003; DeBarry *et al.*, 2006; Hawkins *et al.*, 2006; Ungerer *et al.*, 2006; Feschotte, 2008).

The class 1 elements known as long terminal repeat (LTR) retrotransposons are the most abundant transposable elements in plants and are of particular interest in studies of plant genome size evolution because of their

large size and potential for massive increases in copy number (Kumar and Bennetzen, 1999). Autonomous LTR retrotransposons consist of two superfamilies known as *Ty3/gypsy*-like and *Ty1/copia*-like. While bearing similar structural features overall (for example LTRs in direct orientation, presence of *GAG* and *POL* genes), *Ty3/gypsy* and *Ty1/copia* retrotransposon superfamilies represent separate ancient lineages, which are diverged at the sequence level and also differ with regard to the order of domains within *POL* (Kumar and Bennetzen, 1999). Their 'copy-and-paste' replication mode allows numerous copies of intact elements to be synthesized, reverse transcribed, and integrated into new positions in the genome (Sabot and Schulman, 2006). Additional LTR retrotransposons such as Terminal-repeat Retrotransposons In Miniature (Witte *et al.*, 2001) and Large Retrotransposon Derivatives (Kalendar *et al.*, 2004) also exist in plant genomes. These additional types of LTR retrotransposons lack the capacity for independent transposition and thus are classified as non-autonomous. Differential accumulation (and loss) of LTR retrotransposons in a species-specific manner can contribute substantially to interspecific variation in nuclear DNA content and explain how closely related species can have markedly different genome sizes despite otherwise similar evolutionary histories.

LTR retrotransposons have been found in virtually all plant species in which they have been sought (Voytas *et al.*, 1992; Flavell *et al.*, 1992b; Suoniemi *et al.*, 1998). It has been suggested that their transcriptional and transposi-

Correspondence: Dr MC Ungerer, Division of Biology, Kansas State University, Ackert Hall, Manhattan, KS 66506, USA.

E-mail: mcungere@ksu.edu

Received 19 August 2009; revised 28 October 2009; accepted 9 November 2009; published online 13 January 2010

tional activity is under precise control in host plant genomes because immediate consequences of uncontrolled activation of transposable elements are thought to be deleterious. Transcriptional gene silencing through DNA methylation and chromatin modification is one of the proposed mechanisms that specifically targets transposable element activity in plant genomes (Feschotte *et al.*, 2002). Despite the ubiquity and abundance of these elements, however, they are infrequently found to be transpositionally active, and the environmental and genomic conditions that facilitate their activation and proliferation are not well understood. Two natural phenomena thought to be associated with retrotransposon derepression and proliferation include hybridization and biotic/abiotic stress (Wessler, 1996; Grandbastien, 1998; Waugh O'Neill *et al.*, 1998; Labrador *et al.*, 1999; Liu and Wendel, 2000). Activation following the former may be associated with a genome-wide pattern of hypomethylation that is sometimes observed in the genomes of hybrids (Waugh O'Neill *et al.*, 1998). The mechanistic basis of stress-induced activation may be related to regulatory sequences found in the 5' LTR of retrotransposons that bear strong resemblance to the promoters of certain stress response endogenous plant genes (Grandbastien *et al.*, 1997; Takeda *et al.*, 1999).

Although mounting evidence suggests that hybridization and stress may be important triggers of retrotransposon derepression and proliferation, documenting historical bouts of these phenomena in natural populations and determining whether they were followed by episodes of retrotransposon proliferation is difficult because of the rapid and ephemeral nature of proliferation events. Therefore, identifying examples of recent proliferation events in systems with detailed evolutionary histories may provide an excellent opportunity to further our understanding of these dynamics. Such systems additionally can provide excellent opportunities to examine the consequences of transposable element proliferation on genome function.

A species group that meets many of these criteria includes three diploid hybrid sunflower species from

North America. These species, *Helianthus anomalous*, *H. deserticola*, and *H. paradoxus*, are independently derived via hybridization events between the same two parental taxa, *H. annuus* and *H. petiolaris* (Rieseberg, 1997; Lai *et al.*, 2005). The three hybrid taxa occupy habitats considered abiotically extreme relative to either parental species (Rieseberg *et al.*, 2003). *H. anomalous* and *H. deserticola* inhabit arid desert-like environments, whereas *H. paradoxus* occurs exclusively in saline environments, such as brackish salt marshes and salt seeps. Estimates of nuclear DNA content using flow cytometry suggest that genome sizes of the hybrid taxa are at least 50% larger than those of parental species (Baack *et al.*, 2005). Variation in genome size among these sunflower species is not attributable to polyploidization because all species are diploid and possess the same number of chromosomes ($n=17$). It was shown earlier that *Ty3/gypsy*-like retrotransposons have undergone massive proliferation events in the genomes of all three hybrid taxa and that these proliferation events have contributed substantially to increases in genome size in the hybrid species (Ungerer *et al.*, 2006, 2009; Staton *et al.*, 2009). In this report, we investigate whether patterns of proliferation are evident for the other major superfamily of autonomous retrotransposon (*Ty1/copia*-like retrotransposons) in these sunflower hybrid species. We show that *Ty1/copia*-like elements also have increased in copy number in the sunflower hybrid species, but the scale of proliferation is lower overall (as compared with *Ty3/gypsy*) and different among the three species, with one hybrid species having experienced a relatively large-scale proliferation event of *Ty1/copia* elements and the remaining two species having experienced considerably smaller-scale events.

Materials and methods

Specimens

Seeds of all species under investigation were obtained from the USDA National Plant Germplasm System (Table 1). Seeds were germinated in the dark on moist

Table 1 Source of plant material and summary statistics of the RT-domain-encoding-region (229 bp) of *Ty1/copia*-like elements isolated from five *Helianthus* species

Species	Accession ^a	Original collection location		Number of sequences			S ^e	h ^f	Nucleotide diversity			
		State	Lat. (N), Long. (W)	Total ^b	Indels/ stop ^c	Full length ^d			Total		Sub-lineage B'	
									π	s.d. (π)	π	s.d. (π)
<i>H. annuus</i>	PI 468607	Utah	37.239, 113.358	39	10	29	196	38	0.314	0.034	0.078	0.010
	PI 468477	Texas	35.192, 102.107	—	—	—	—	—	—	—	—	—
<i>H. petiolaris</i>	PI 468815	Utah	37.047, 112.530	48	10	38	191	46	0.194	0.033	0.077	0.008
	PI 435807	Texas	35.190, 102.065	—	—	—	—	—	—	—	—	—
<i>H. anomalous</i>	Ames 26095	Utah	39.744, 112.316	47	16	35	182	47	0.235	0.036	0.073	0.010
<i>H. deserticola</i>	Ames 26094	Utah	37.254, 113.343	46	21	27	192	43	0.278	0.033	0.083	0.012
<i>H. paradoxus</i>	PI 468802	Texas	30.883, 102.983	41	9	35	183	39	0.200	0.038	0.049	0.004

Abbreviations: PCR, polymerase chain reaction; RT, reverse transcriptase.

Nucleotide diversity (π) was calculated for all sequences within each species' genome as well as for the sub-lineage B' (see Figure 3). Accession PI 468477 (*H. annuus*) and accession PI 435807 (*H. petiolaris*) were used only in the quantitative PCR assays of element copy number.

^aUSDA-National Plant Germplasm System (<http://www.ars-grin.gov/npgs/>).

^bThe total number of sequences obtained.

^cThe number of sequences with indels or putative stop codons.

^dThe number of sequences without putative stop codons.

^eThe number of segregating sites.

^fThe number of unique sequences.

filter paper in Petri dishes, and germinated seedlings were then transferred to four-inch pots and grown in the Kansas State University greenhouses until suitable size for harvesting of plant tissue. DNA isolated from a single plant of each species was used in Southern blots and for assessing *Ty1/copia* sequence heterogeneity via degenerate polymerase chain reaction (PCR) accompanied by clone sequencing, while the following numbers of individuals were used for quantitative PCR assays of element numerical abundance in the individual species' genomes: *H. annuus* (Utah), $n=10$; *H. annuus* (Texas), $n=10$; *H. anomalus*, $n=10$; *H. deserticola*, $n=10$; *H. paradoxus*, $n=6$; *H. petiolaris* (Utah), $n=9$; *H. petiolaris* (Texas), $n=10$. DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) following the manufacturer's instructions.

Southern blots

Genomic digests were performed with the restriction enzyme *Hae*III (Promega, Madison WI, USA), separated on 0.8% agarose, and stained with ethidium bromide before capillary transfer to a Hybond-N+ membrane (Amersham Biosciences, Buckinghamshire, UK). The amount of DNA for restriction digestion was standardized both by genome equivalents and by mass (1 μ g). For the genome equivalent standardization, $\sim 1.5 \times 10^5$ nucleus equivalents was estimated for each of the five species based on published (Baack *et al.*, 2005) 2C values (*H. annuus*, 1.10 μ g; *H. anomalus*, 1.76 μ g; *H. deserticola*, 1.63 μ g; *H. paradoxus*, 1.63 μ g; *H. petiolaris*, 1.00 μ g). Two PCR probes were used separately in Southern blots: a 230-bp region of the *Ty1/copia* ribonuclease H (RNase H) domain amplified with non-degenerate primers and a 263-bp region of the *Ty1/copia* reverse transcriptase (RT) domain amplified with degenerate primers. These regions were amplified from genomic DNA of *H. annuus*, a parental species, with the following primers: forward, 5'-TCTCAGAACCCTCGGCAATCT-3', and reverse, 5'-GGCGAGCAAAGAGAAAATG-3', for RNase H domain; and forward, 5'-ACNGCNTTYTNCAYGG-3', and reverse, 5'-ARCATRTCRTCNCACRTA-3', for RT domain. Primers targeting the RNase H domain were designed based on a sequence reported previously in sunflower (GenBank accession no. AJ532591). Degenerate primers targeting the RT domain were developed earlier (Flavell and Smith, 1992; Flavell *et al.*, 1992a,b). For probe preparation, standard PCR amplifications included 1 \times reaction buffer (Promega), 0.2 mM of each dNTP, 5 pmol of each primer, 1 U of GoTaq Flexi (Promega), 2 mM MgCl₂, and 2.5 μ l of template DNA (c. 50 ng) in a 25- μ l reaction volume. PCR was performed on a MJ Research (Watertown, MA, USA) PTC-100 thermal cycler under the following conditions: 3 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 56 °C (RNase H probe amplification) or 45 °C (RT probe amplification), and 60 s at 72 °C, and a final extension of 3 min at 72 °C. PCR amplifications of these probes yielded a single band of expected size on an agarose gel. Sequencing of these probes confirmed that they were *Ty1/copia* RNase H and RT fragments, respectively. The PCR probes were purified using a QIAquick PCR Purification Kit (Qiagen) before labelling with 32-P dCTP using a Prime-It RmT Random Primer Labeling Kit (Stratagene, La Jolla, CA, USA). Hybridization was performed overnight at 65 °C

in a solution consisting of 6 \times SSC, 10 \times Denhardt's reagent, 25 mM NaPB, 0.5% SDS, 10% PEG, 200 μ g ml⁻¹ salmon sperm denatured by boiling, and labelled probe. Following hybridization, membranes were washed at 65 °C with 2 \times SSC, 0.1% SDS (2 \times 15 min) followed by additional washes at 65 °C with 0.2 \times SSC, 0.1% SDS (2 \times 30 min). Southern blot images were captured via a STORM840 phosphoimaging system (Molecular Dynamics, Sunnyvale, CA, USA).

Quantitative PCR assay

Quantitative PCR (Q-PCR) assays of element numerical abundance were conducted using an iCycler iQ quantitative PCR system (Bio-Rad, Hercules, CA, USA) with initial denaturing at 94 °C for 90 s, followed by 40 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s. Reactions were performed using the iQ SYBR Green Supermix kit (Bio-Rad) following the manufacturer's protocols, with the exception that reactions were conducted in 25 μ l volumes. All reactions were performed using 1 μ l of template DNA (standardized to 80 pg μ l⁻¹) with the same non-degenerate primer pair used to generate a probe of the RNase H fragment for Southern blots.

Copy numbers of *Ty1/copia* RNase H sequences in the hybrid and parental species' genomes were estimated following a standard curve method described in Ungerer *et al.* (2006) and based on protocols described in Bustin (2000). Briefly, a standard curve was generated by cloning the 230-bp *Ty1/copia* RNase H fragment (from *H. annuus*), determining the concentration of isolated plasmid (and insert) via a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA), estimating plasmid number per unit volume, and determining Ct values for a dilution series of the cloned *Ty1/copia* RNase H fragment at 5.0, 0.5, 0.05, 0.005, and 0.0005 ng μ l⁻¹. Dilution series were performed in duplicate. PCR efficiency for the standard curve exceeded 96%. Copy number estimates were converted to copy number per genome using mean C values for each species estimated from several populations (Baack *et al.*, 2005). Data from two independent Q-PCR runs were combined to produce Figure 2. Copy number estimates (per genome) were compared among the five species using non-parametric Kruskal-Wallis test in the software JMP 7.0.1 (SAS institute, Cary, NC, USA).

Ty1/copia sequence diversity survey and phylogenetic analysis

The same degenerate primers used to amplify the RT probe used in Southern hybridization experiments were used to assess sequence variation of *Ty1/copia* elements in these sunflower species. For each species, five individual 25 μ l PCR reactions were performed and pooled for further processing to reduce potential effects of PCR drift (Wagner *et al.*, 1994). Pooled products were gel purified using a QIAquick Gel Extraction Kit (Qiagen) and cloned using the pGEM-T Vector System I (Promega). Between 39 and 48 cloned PCR products per species were sequenced using the M13 universal sequencing primer on an ABI 3730xl Genetic Analyzer (Foster City, CA, USA). Average nucleotide diversity (π) was calculated using DnaSP version 5.00.03 (Rozas *et al.*, 2003).

Sequence alignments were conducted with ClustalW (Thompson *et al.*, 1994) with subsequent manual adjust-

ments. Removal of primer sequences resulted in reads of ~229 bp. Phylogenetic analyses were conducted using neighbour joining (NJ) and maximum parsimony (MP) in PAUP* 4.0b10 (Swofford, 2000). Genetic distances for NJ were calculated using a maximum-likelihood model of sequence evolution. The best-fit model of sequence evolution was assessed using the Akaike information criterion in ModelTest version 3.7 (Posada and Crandall, 1998). The GTR+I+G (six substitution types = 1.97, 4.43, 1.64, 1.07, 5.76, and 1.00; I = 0.032; G = 0.991) was inferred as the best-fit model for our dataset. For MP analyses, an unweighted heuristic search was performed with tree bisection reconnection branch swapping, character state optimization using the MINF option, and random addition of taxa with 10 replications. The reliability of tree topologies from these analyses was estimated using the bootstrap with 1000 pseudoreplicates. NJ and MP analyses produced near-identical topologies. We thus present results based on NJ only, but include bootstrap support levels for both NJ and MP analyses. Sequences used in this study have been deposited in GenBank (Accession nos. GU108960–GU109180).

Both NJ and MP analyses revealed instances of large numbers of closely related sequences. As multifurcating network analyses better represent complex evolutionary relationships among closely related sequences than strictly bifurcating phylogenetic trees (Posada and Crandall, 2001), we used median-joining (MJ) networks of Ty1/copia-like RT-domain in follow-up analyses of particular sets of sequences (sub-lineage B' in Figures 3b and 4). MJ networks were constructed using the software Network version 4.510 (<http://www.fluxus-engineering.com>) (Bandelt et al., 1995) with equal weights to all variable sites, epsilon parameter = 0, and the MP option to remove non-MP links from the network.

Evolutionary relationships of Ty1/copia-like elements in *Helianthus* relative to other plant Ty1/copia-like retrotransposons were evaluated by phylogenetic analysis of aligned amino-acid sequences of the RT domain. Amino-acid sequences ($n = 142$) of sunflower Ty1/copia-like RT domains lacking premature stop codons and/or apparent frameshift mutations were aligned with the RT regions of 18 plant Ty1/copia-like retrotransposons reported in Peterson-Burch and Voytas (2002), an earlier published sequence isolated from *H. annuus* (GenBank accession no. AJ532591), and the RT region of a recently isolated full-length Ty1/copia-like element (*HACRE1*) also isolated from *H. annuus* (Buti et al., 2009; EMBL accession no. FN298618). Alignments were performed with ClustalW and were subsequently adjusted manually with the aid of retroelement alignments reported in Xiong and Eickbush (1990). The *Athila* (*Arabidopsis*) and *RIRE3* (*Oryza*) Ty3/gypsy-like elements were used as phylogenetic outgroups (GenBank Accession nos. AB005248 and AB014738, respectively). An NJ phylogeny with Poisson correction was generated using MEGA version 4.0.2 (Tamura et al., 2007) with 1000 bootstrap pseudoreplicates for assessing the reliability of tree topology. An MP phylogeny was generated using an unweighted heuristic search, tree bisection reconnection branch swapping, and MINF option, in PAUP* 4.0b10 with 1000 bootstrap pseudoreplicates. NJ and MP analyses produced near-identical topologies. We thus present results based on NJ only, but include bootstrap support levels for both NJ and MP analyses.

Results

Ty1/copia-like retrotransposon abundance in parental and hybrid species' genomes

Relative abundance of Ty1/copia-like LTR retrotransposons in the genomes of three hybrid sunflower species (*H. anomalus*, *H. deserticola*, and *H. paradoxus*) and their parental species (*H. annuus* and *H. petiolaris*) was compared using Southern blots and quantitative PCR. Southern blots using a PCR generated probe of the RNase H-domain-encoding region exhibited similar hybridization signal intensity among the five species investigated, with the exception of *H. paradoxus*, for which hybridization signal intensity was considerably stronger (Figure 1). This pattern was especially noticeable when loadings were standardized by genome equivalents. Stronger hybridization signal suggests a higher relative abundance of Ty1/copia-like sequences in the *H. paradoxus* genome as compared with the genomes of the other species. Stronger hybridization signal in *H. paradoxus* was less detectable in Southern blots using a PCR probe of the RT-domain-encoding region (Supplementary Figure S1). As the RT probe was amplified with degenerate primers, a more diverse array of sequences (see next section) served as probes during hybridization; hybridization-based differences in signal intensity attributable to proliferation of a single sub-lineage of Ty1/copia was thus likely masked.

Estimates of element copy number per genome as determined from quantitative PCR assays are depicted in Figure 2 for the three hybrid species and for two different populations each of the parental species. Copy number estimates of the Ty1/copia RNase H domain were significantly different among the seven populations (Kruskal–Wallis test, d.f. = 6, $\chi^2 = 37.58$, $P < 0.0001$) (Figure 2). Consistent with results of Southern blot

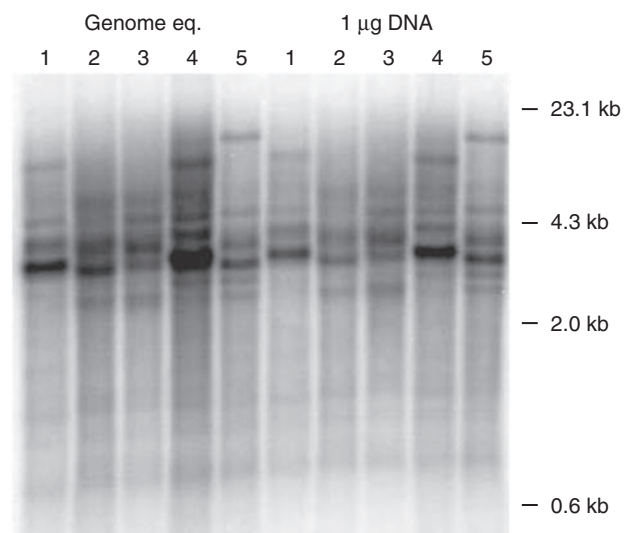


Figure 1 Southern blots of Ty1/copia RNase H domain hybridized to the genomic DNAs of two parental and three hybrid sunflower species. Lanes 1 and 5: parental species *H. annuus* and *H. petiolaris*, respectively. Lanes 2–4: hybrid species *H. anomalus*, *H. deserticola*, and *H. paradoxus*, respectively. Amount of genomic DNAs for loading was standardized by either genome equivalent (left five lanes) or 1 µg of DNA (right five lanes).

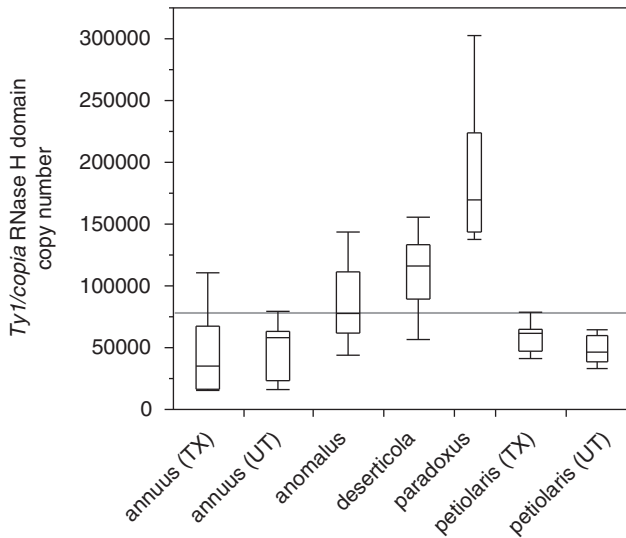


Figure 2 Copy number estimates (per genome) of the *Ty1/copia* RNase H domain based on quantitative PCR. Two different populations of the *Helianthus* parental species (*H. annuus* and *H. petiolaris*) and one population each of the three diploid hybrid species (*H. anomalus*, *H. deserticola*, and *H. paradoxus*) were assayed. The number of individuals assayed per species/population are as follows: *H. annuus* (Utah), $n=10$; *H. annuus* (Texas), $n=10$; *H. anomalus*, $n=10$; *H. deserticola*, $n=10$; *H. paradoxus*, $n=6$; *H. petiolaris* (Utah), $n=9$; *H. petiolaris* (Texas), $n=10$. The horizontal line across box plots indicates the global mean of copy number estimates among the five species.

analyses with an RNase H probe, *H. paradoxus* had the highest copy number estimate (3.7-fold increase in copy number relative to parental species). The Q-PCR assay additionally indicates that element copy numbers also are elevated in the other two hybrid species relative to the parental species, albeit to a lesser scale, with increases in copy number of 1.7-fold and 2.2-fold relative to the parental species for *H. anomalus* and *H. deserticola*, respectively.

Diversity of *Ty1/copia* in the hybrid and parental sunflower species

Sequence diversity of *Ty1/copia*-like LTR retrotransposons was surveyed in the three hybrid and two parental sunflower species by amplifying a 229-bp (excluding PCR primers) RT-domain-encoding fragment with degenerate primers (see Materials and methods) followed by cloning and sequencing 39–48 PCR products per species. Analysis of these sequences revealed considerable diversity in each of the five species (Table 1). From 182 to 196 nucleotide sites of the RT fragment showed variation, and from 38 to 47 unique sequences were identified in each species. Within species, nucleotide diversity (π) of the 229-bp RT fragment ranged from 0.194 to 0.314. From nine to 21 insertion–deletion polymorphisms and premature stop codons were identified across the five species, suggesting that several of these *Ty1/copia*-like elements have lost the ability to transpose autonomously.

Phylogenetic analyses of these sequences resolved several lineages with moderate-to-strong bootstrap support (Figure 3). Sequences from the five sunflower taxa were distributed across the entire tree (as opposed to

species-specific lineages of *Ty1/copia*-like sequences), indicating that these *Ty1/copia*-like lineages are ancient and predate the divergence of the parental species *H. annuus* and *H. petiolaris*. The majority (70%) of sequences were concentrated in a single sub-lineage (labelled B' in Figure 3). An MJ network for sub-lineage B' identified three genetic clusters (Figure 4). Cluster I was the largest, and composed of sequences derived from all five species, whereas Clusters II and III were composed of sequences derived from all species except *H. paradoxus*.

Phylogenetic relationship to other plant *Ty1/copia*-like retrotransposons

A phylogenetic analysis of 142 translated sequences from this study coupled with a diverse set of previously characterized *Ty1/copia*-like sequences from other plants is depicted in Figure 5, with terminal branches of *Helianthus* *Ty1/copia*-like amino-acid sequences from this study indicated in red. Sequences from sunflower *Ty1/copia* lineage B (see Figure 3), which were most abundant in the *Helianthus* genome, form a single, exclusive cluster, which includes an additional previously published sequence isolated from *H. annuus* (AJ532591). However, other RT sequences isolated from *Helianthus* species were distributed more broadly across the tree, indicating that they are more closely related to elements isolated from the genomes of other plant species and thus are members of more ancient *Ty1/copia*-like lineages that predate the species divergences of the host taxa.

Discussion

In earlier reports (Ungerer *et al.*, 2006, 2009), we showed that the genomes of these diploid hybrid sunflower species experienced proliferation events of a major superfamily of LTR retrotransposon, the *Ty3/gypsy*-like group, following, or associated with their origins. Noteworthy of those studies was that proliferation of *Ty3/gypsy*-like elements was of similar scale among the three hybrid sunflower taxa with regard to copy number increases and that the same element sub-lineage served as the proliferative source lineage. In this report, we examine whether similar derepression and amplification events occurred for the other major superfamily of autonomous LTR retrotransposons, the *Ty1/copia*-like group. Two independent assays of copy number abundance (Southern blots and Q-PCR) revealed evidence of derepression and amplification of *Ty1/copia*-like sequences in the genome of *H. paradoxus*; Q-PCR assays additionally suggest amplification of these elements in the other two hybrid species, albeit to a lesser scale.

Copy number increases of both *Ty3/gypsy*-like and *Ty1/copia*-like elements in hybrid sunflower taxa imply that genetic mechanisms to repress LTR retrotransposon activity may not be element superfamily specific but are more likely to be general control mechanisms affecting different classes of transposable elements simultaneously throughout the genome. Such mechanisms may include transcriptional gene silencing through DNA methylation and chromatin modification, the disruption of which may represent a form of 'genome shock' (McClintock, 1984) that may be facilitated by hybridization and/or environmental stress. Interestingly,

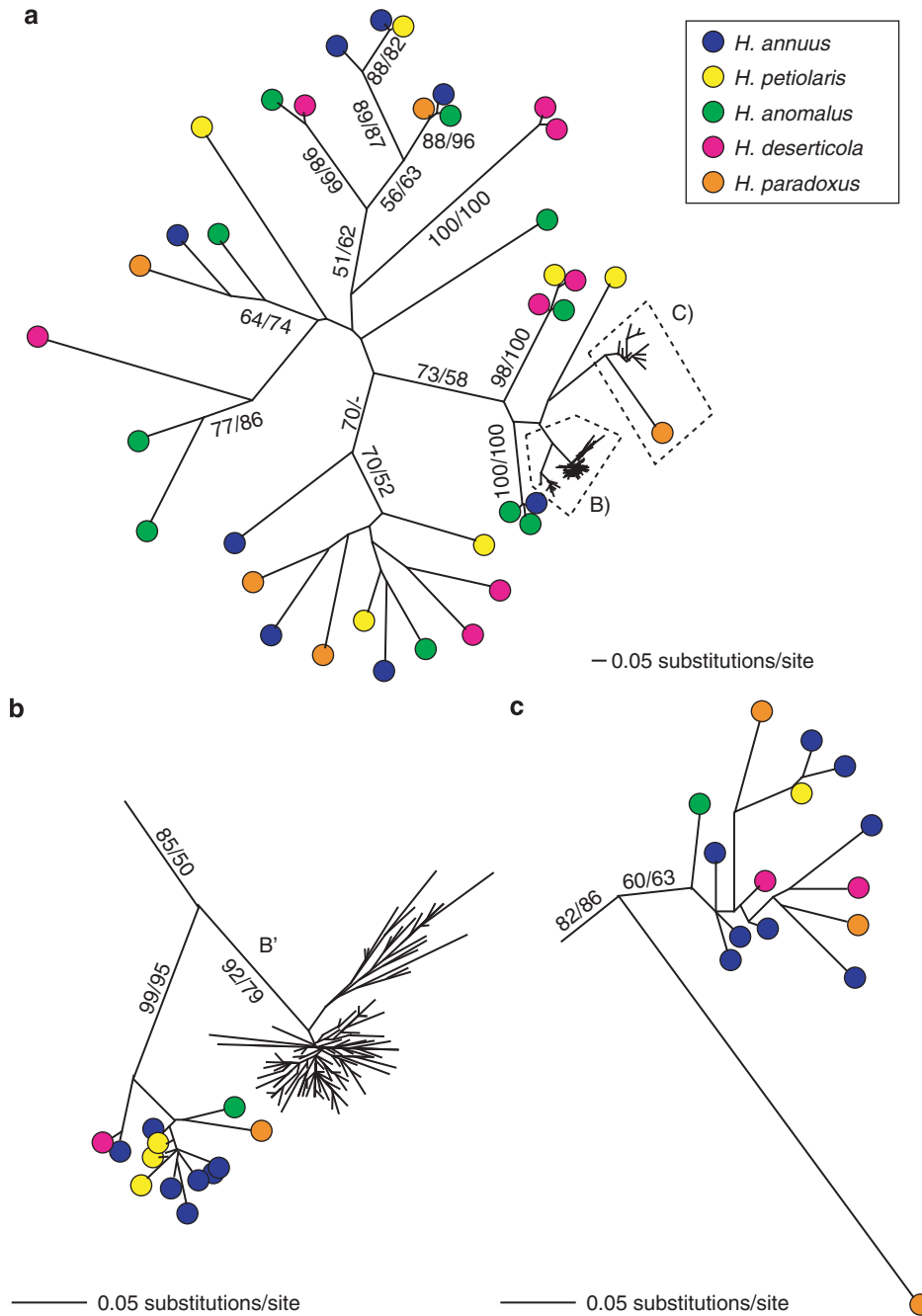


Figure 3 (a) Neighbour-joining phylogram based on a 229-bp region of the *Ty1/copia*-like reverse transcriptase (RT) domain for the five *Helianthus* species (*H. annuus*, *H. petiolaris*, *H. anomalus*, *H. deserticola*, and *H. paradoxus*). Numbers on branches indicate bootstrap supports (NJ/MP). Enlarged lineages B and C are shown in (b) and (c). Sub-lineage B' is shown in Figure 4. Different colours depict different *Helianthus* species.

however, the overall increases in copy number of *Ty1/copia*-like elements are far lower than that observed for *Ty3/gypsy*-like elements, and additionally, the scale of copy number increase differs considerably among the three sunflower hybrid species, with a 3.7-fold increase in copy number in the genome of *H. paradoxus* (relative to the average parental species value) versus a lower 1.7-fold and 2.2-fold increase for *H. anomalus* and *H. deserticola*, respectively. In detecting differences among species in element abundance, assays performed by quantitative PCR were considerably more sensitive

than comparisons of hybridization signal intensity by Southern blot analysis.

Differences among the hybrid species in the scale of proliferation (that is copy number increases) may reflect random dynamics of *Ty1/copia*-like element derepression in the hybrid species during or following their independent origins. As a possible alternative explanation, however, it is interesting to note that the habitat of *H. paradoxus* differs markedly from that of the other two hybrid taxa. *H. paradoxus* is found in saline environments, whereas *H. anomalus* and *H. deserticola* are found

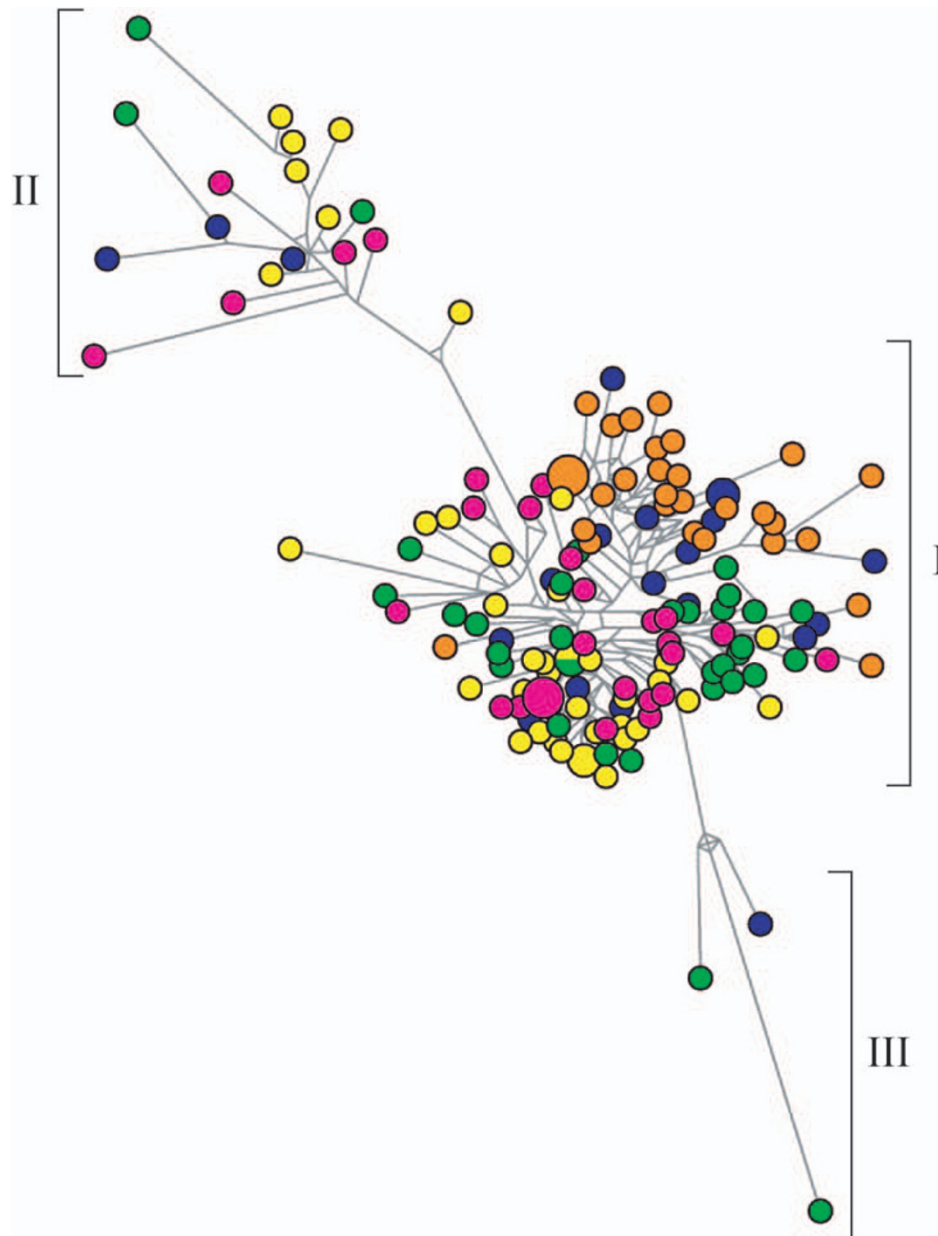


Figure 4 Median-joining network for sub-lineage B' of *Ty1/copia*-like RT domain. Line lengths and node area are proportional to branch lengths and number of sequences, respectively. Three clusters are indicated (I–III). Species colour codes are the same as in Figure 3.

in more desert-like habitats. Environmental stress has been implicated as a causal agent of LTR retrotransposon derepression and proliferation in plants. It thus seems conceivable that the different habitats and associated environmental stresses experienced by the hybrid species could be associated with these different scales of proliferation, especially given that exposure of plants to high salt concentrations has been shown to induce *Ty1/copia*-like transcriptional activity in other species (Tapia *et al.*, 2005; De Felice *et al.*, 2009).

A survey of *Ty1/copia*-like diversity (based on the RT region) within the genomes of these sunflower taxa using degenerate PCR followed by sequencing of multiple clones revealed considerable element heterogeneity, with sequence diversity estimates (π) ranging from 0.194 to 0.314 (Table 1). The high level of diversity of *Ty1/copia*-like elements sampled from the genomes of these species

is confirmed through phylogenetic analysis of these sequences with elements isolated from other plant species (Figure 5). Sequences isolated from sunflower did not exhibit exclusive monophyly but rather were found distributed throughout the evolutionary tree (Figure 5). Given the taxonomic breadth of the host species used in this analysis (11 genera from 5 families and including both monocots and dicots), it is clear that several ancient and diverse lineages of *Ty1/copia*-like elements exist in the sunflower genome, a situation considered normal for plant species (Flavell *et al.*, 1992b; Marin and Llorens, 2000; Zhang and Wessler, 2004).

Although considerable sequence diversity was revealed in the survey of *Ty1/copia*-like elements, phylogenetic analysis of sequences from *Helianthus* species alone (Figures 3 and 4) show that a majority of sequences (70%) belong to a single sub-lineage, designated as

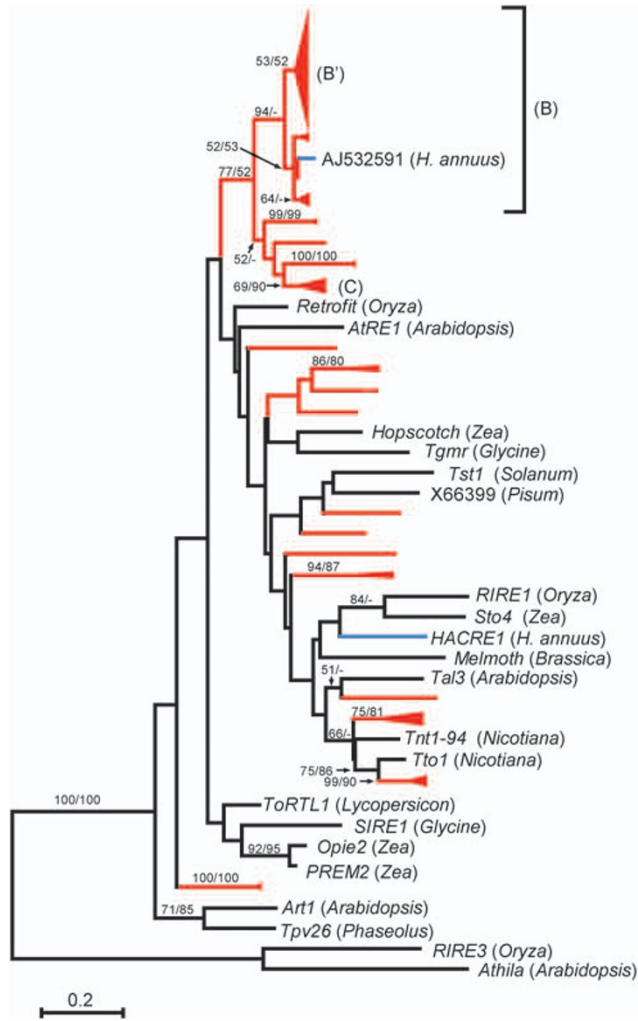


Figure 5 Neighbour-joining phylogram based on RT amino-acid sequences of *Ty1/copia*-like elements isolated from *Helianthus* and other plants. Terminal branches shown in red represent elements isolated from *Helianthus* species (this study), and those in blue represent elements previously isolated from *H. annuus* by other researchers (see Materials and methods). Clades composed exclusively of sequences identified in this study are compressed and represented as red triangles at the terminal ends of branches. Numbers on branches indicate bootstrap support (NJ/MP). B, B', and C indicate lineages/sub-lineages identified in Figure 3. Sequences from this study that possess premature stop codons and/or frameshift mutations were excluded from the analysis.

lineage B' in Figure 3. Sequences from within this sub-lineage were highly similar, as indicated by short branch lengths and lower estimates of sequence diversity (π) as compared with analyses of all isolated sequences (Table 1). Groups of sequences with short branch lengths and low values of pairwise sequence divergence are consistent with predictions for retrotransposon lineages that proliferated in recent evolutionary time, given that proliferative elements give rise to daughter copies that are identical at the time of insertion and subsequently acquire mutations independently. Further analysis of this sub-lineage by MJ analysis revealed that it is composed of three sequence clusters, one of which harbours a majority of the sequences isolated (Cluster I in Figure 4) and may likely represent the source lineage of proliferation of *Ty1/copia*-like elements in the hybrid taxa.

Additional evidence supporting this lineage as the proliferative lineage in the hybrid species' genomes is obtained by examining the position of a *Ty1/copia*-like sequence previously isolated from *H. annuus* (GenBank accession no. AJ532591) within the NJ phylogram depicted in Figure 5. This 1663-bp sequence encompasses adjacent RT and RNase H domains of a single, putative *Ty1/copia*-like retrotransposon within the *H. annuus* genome. In addition to its inclusion in the phylogenetic analysis shown in Figure 5, this sequence also was used to design primers targeting the RNase H domain for assays of element numerical abundance via Southern blotting and Q-PCR (see Figures 1 and 2). Thus, the finding of elevated copy numbers of this RNase H domain variant in the hybrid species' genomes is fully consistent with its phylogenetic placement (based on the RT domain) within a lineage exhibiting a signal of recent amplification based on shorter branch lengths and lower pairwise sequence divergence values. This pattern is especially evident for *H. paradoxus*, for which copy number of *Ty1/copia*-like elements was highest (Figure 2) and nucleotide diversity (π) within sub-lineage B' was correspondingly lowest (Table 1).

Although such patterns are exactly those predicted under a scenario of transposable element proliferation in a host genome, a few caveats deserve mention with regard to our diversity survey. First, the heterogeneity of sequences recovered in this study may be limited by the degeneracy of the primers used. Although a potential limitation, the pool of sequences recovered in our survey is remarkably diverse, with *Helianthus* sequences dispersed throughout the phylogenetic tree of *Ty1/copia*-like elements isolated from a diverse group of plants (Figure 5). Second, although we cannot definitively exclude PCR recombination (Meyerhans *et al.*, 1990; Yang *et al.*, 1996; Bradley and Hillis, 1997) or other artefacts as having artificially contributed to observed levels of sequence diversity in this study, the consequences of such phenomena, if present at all in our data, are unlikely to bias our major conclusions given our assertion that the candidate proliferative source lineage is characterized by a large number of sequences exhibiting low levels of nucleotide diversity. Finally, a third caveat is that although the progenitor species that gave rise to the hybrid taxa remain extant, certain *Ty1/copia*-like retrotransposon lineages may have been lost from the progenitor species' genomes over evolutionary time, and thus the genomic composition of modern day *H. annuus* and *H. petiolaris* may differ from that of the actual *H. annuus* and *H. petiolaris* individuals/populations from which the hybrid taxa were derived. It is unlikely that differences in genome size between the hybrid species and their parental taxa are due exclusively to genomic downsizing in the parent species, however, given that genome size in the three hybrid species is considerably elevated in comparison with all *Helianthus* diploid annual species, not just *H. annuus* and *H. petiolaris* (Sims and Price, 1985).

Conclusion

The genomic and/or environmental conditions during the origins of these hybrid sunflower taxa were conducive to derepression and amplification of two superfamilies of LTR retrotransposon. The scale of

proliferation, however, was far greater for *Ty3/gypsy*-like elements than for *Ty1/copia*-like elements. Differences presumably relate to triggers of transcriptional/transpositional activation and counterbalancing mechanisms of repression by the host genome. We are currently exploring the extent to which these elements may be activated in natural hybrid zones and in early generation greenhouse hybrids derived from the parental species *H. annuus* and *H. petiolaris*. The extent to which other Class I transposable elements, such as Long Interspersed Nuclear Elements and Large Retrotransposon Derivatives, may have become active and proliferated in the hybrid species' genomes has not yet been determined.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dr Lucia Natali for providing us with the *HACRE1* sequence before publication and three anonymous reviewers for comments that improved this manuscript. This work was supported by Kansas State University and National Science Foundation grant DEB-0742993 to MCU.

References

Baack EJ, Whitney KD, Rieseberg LH (2005). Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytol* **167**: 623–630.

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995). Mitochondrial portraits of human-populations using median networks. *Genetics* **141**: 743–753.

Bennetzen JL (2000). Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* **42**: 251–269.

Bowen NJ, Jordan IK (2002). Transposable elements and the evolution of eukaryotic complexity. *Curr Issues Mol Biol* **4**: 65–76.

Bradley RD, Hillis DM (1997). Recombinant DNA sequences generated by PCR amplification. *Mol Biol Evol* **14**: 592–593.

Bustin SA (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**: 169–193.

Buti M, Giordani T, Vukich M, Gentzbittel L, Pistelli L, Cattonaro F et al. (2009). *HACRE1*, a recently inserted *copia*-like retrotransposon of sunflower (*Helianthus annuus* L.). *Genome* **52**: 904–911.

De Felice B, Wilson RR, Argenziano C, Kafantaris I, Conicella C (2009). A transcriptionally active *copia*-like retroelement in *Citrus limon*. *Cell Mol Biol Lett* **14**: 289–304.

DeBarry JD, Ganko EW, McCarthy EM, McDonald JF (2006). The contribution of LTR retrotransposon sequences to gene evolution in *Mus musculus*. *Mol Biol Evol* **23**: 479–481.

Feschotte C (2008). Opinion—transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9**: 397–405.

Feschotte C, Jiang N, Wessler SR (2002). Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* **3**: 329–341.

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992a). *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res* **20**: 3639–3644.

Flavell AJ, Smith DB (1992). A *Ty1-copia* group retrotransposon sequence in a vertebrate. *Mol Gen Genet* **233**: 322–326.

Flavell AJ, Smith DB, Kumar A (1992b). Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. *Mol Gen Genet* **231**: 233–242.

Grandbastien MA (1998). Activation of plant retrotransposons under stress conditions. *Trends Plant Sci* **3**: 181–187.

Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta JM (1997). The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. *Genetica* **100**: 241–252.

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res* **16**: 1252–1261.

Kalendar R, Vicient CM, Peleg O, Ananthawat-Jonsson K, Bolshoy A, Schulman AH (2004). Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. *Genetics* **166**: 1437–1450.

Kidwell MG, Lisch D (1997). Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA* **94**: 7704–7711.

Kumar A, Bennetzen JL (1999). Plant retrotransposons. *Annu Rev Genet* **33**: 479–532.

Labrador M, Farre M, Utzet F, Fontdevila A (1999). Interspecific hybridization increases transposition rates of *Oswaldo*. *Mol Biol Evol* **16**: 931–937.

Lai Z, Nakazato T, Salmaso M, Burke JM, Tang S, Knapp SJ et al. (2005). Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* **171**: 291–303.

Liu B, Wendel JF (2000). Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome* **43**: 874–880.

Marin I, Llorens C (2000). *Ty3/Gypsy* retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. *Mol Biol Evol* **17**: 1040–1049.

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

McDonald JF (1995). Transposable elements: possible catalysts of organismic evolution. *Trends Ecol Evol* **10**: 123–126.

Meyerhans A, Vartanian JP, Wainhobson S (1990). DNA recombination during Pcr. *Nucleic Acids Res* **18**: 1687–1691.

Miller WJ, McDonald JF, Nouaud D, Anxolabehere D (1999). Molecular domestication—more than a sporadic episode in evolution. *Genetica* **107**: 197–207.

Peterson-Burch BD, Voytas DF (2002). Genes of the Pseudoviridae (*Ty1/copia* retrotransposons). *Mol Biol Evol* **19**: 1832–1845.

Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Posada D, Crandall KA (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* **16**: 37–45.

Rieseberg LH (1997). Hybrid origins of plant species. *Ann Rev Ecol Syst* **28**: 359–389.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–1216.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Sabot F, Schulman AH (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* **97**: 381–388.

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A et al. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.

Sims LE, Price HJ (1985). Nuclear DNA content variation in *Helianthus* (Asteraceae). *Am J Bot* **72**: 1213–1219.

- Staton ES, Ungerer MC, Moore RC (2009). The genomic organization of Ty3/gypsy-like retrotransposons in *Helianthus* (Asteraceae) homoploid hybrid species. *Am J Bot* **96**: 1646–1655.
- Suoniemi A, Tanskanen J, Schulman AH (1998). Gypsy-like retrotransposons are widespread in the plant kingdom. *Plant J* **13**: 699–705.
- Swofford DL (2000). *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods): Version 4*. Sinauer Associates: Sunderland, Massachusetts.
- Takeda S, Sugimoto K, Otsuki H, Hirochika H (1999). A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J* **18**: 383–393.
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) Software Version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Tapia G, Verdugo I, Yanez M, Ahumada I, Theoduloz C, Cordero C *et al.* (2005). Involvement of ethylene in stress-induced expression of the *TLC1.1* retrotransposon from *Lycopersicon chilense* Dun. *Plant Physiol* **138**: 2075–2086.
- Thompson JD, Higgins DG, Gibson TJ (1994). Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Ungerer M, Strakosh S, Stimpson K (2009). Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. *BMC Biol* **7**: 40.
- Ungerer MC, Strakosh SC, Zhen Y (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr Biol* **16**: R872–R873.
- van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* **19**: 530–536.
- Voytas DF, Cummings MP, Koniczny A, Ausubel FM, Rodermel SR (1992). copia-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA* **89**: 7124–7128.
- Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P *et al.* (1994). Surveys of gene families using polymerase chain-reaction—PCR selection and PCR drift. *Syst Biol* **43**: 250–261.
- Waugh O'Neill RJ, O'Neill MJ, Marshall Graves JA (1998). Undermethylation associated with retroelement activation and chromosomal remodelling in an interspecific mammalian hybrid. *Nature* **393**: 68–72.
- Wessler SR (1996). Turned on by stress. Plant retrotransposons. *Curr Biol* **6**: 959–961.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B *et al.* (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**: 973–982.
- Witte CP, Le QH, Bureau T, Kumar A (2001). Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc Natl Acad Sci USA* **98**: 13778–13783.
- Xiong Y, Eickbush TH (1990). Origin and evolution of retroelements based upon their reverse-transcriptase sequences. *EMBO J* **9**: 3353–3362.
- Yang YL, Wang GQ, Dorman K, Kaplan AH (1996). Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. *Aids Res Hum Retrov* **12**: 303–306.
- Zhang X, Wessler SR (2004). Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proc Natl Acad Sci USA* **101**: 5589–5594.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)