Horizontal Transfer of Non-LTR Retrotransposons from Arthropods to Flowering Plants

Dongying Gao,*[,1] Ye Chu,[2] Han Xia,[1,3] Chunming Xu,[1] Karolina Heyduk,[4] Brian Abernathy,[1] Peggy Ozias-Akins,[2] James H. Leebens-Mack,[4] and Scott A. Jackson*[,1]

[1]Center for Applied Genetic Technologies, University of Georgia, Athens, GA

[2]Department of Horticulture, University of Georgia, Tifton, GA

[3]Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Jinan, Shandong, China

[4]Department of Plant Biology, University of Georgia, Athens, GA

***Corresponding authors:** E-mails: sjackson@uga.edu; dgao@uga.edu.

Associate editor: Brandon Gaut

Abstract

Even though lateral movements of transposons across families and even phyla within multicellular eukaryotic kingdoms have been found, little is known about transposon transfer between the kingdoms Animalia and Plantae. We discovered a novel non-LTR retrotransposon, AdLINE3, in a wild peanut species. Sequence comparisons and phylogenetic analyses indicated that AdLINE3 is a member of the RTE clade, originally identified in a nematode and rarely reported in plants. We identified RTE elements in 82 plants, spanning angiosperms to algae, including recently active elements in some flowering plants. RTE elements in flowering plants were likely derived from a single family we refer to as An-RTE. Interestingly, An-RTEs show significant DNA sequence identity with non-LTR retroelements from 42 animals belonging to four phyla. Moreover, the sequence identity of RTEs between two arthropods and two plants was higher than that of homologous genes. Phylogenetic and evolutionary analyses of RTEs from both animals and plants suggest that the An-RTE family was likely transferred horizontally into angiosperms from an ancient aphid(s) or ancestral arthropod(s). Notably, some An-RTEs were recruited as coding sequences of functional genes participating in metabolic or other biochemical processes in plants. This is the first potential example of horizontal transfer of transposons between animals and flowering plants. Our findings help to understand exchanges of genetic material between the kingdom Animalia and Plantae and suggest arthropods likely impacted on plant genome evolution.

Key words: **genome evolution, horizontal transfer, non-LTR retrotransposon, flowering plants, arthropods.**

Article

Introduction

One cornerstone of Mendelian genetics is the transmission of genetic material from parent to offspring, vertical gene transfer (VGT). However, a growing number of studies provide support for the exchange of heritable material between reproductively isolated species, horizontal gene transfer (HGT) (Soucy et al. 2015). Acquisition of foreign DNA may result in traits beneficial to recipients, such as drug and disease resistance (Zhu and Gao 2014). Thus, HGT is viewed as an important force in genome evolution and adaption of both prokaryotes and eukaryotes (Koonin et al. 2001; Keeling and Palmer 2008). It has been estimated that >80% of prokaryotic genes were historically derived from HGT (Dagan et al. 2008). However, HGT in multicellular eukaryotes appears to be far less common than in prokaryotes (Keeling and Palmer 2008).

Transposable elements (TEs) are more prone to horizontal transfer as compared with other DNA sequences, for example, genes, because of their mobility and ability to integrate into chromosomes (Schaack et al. 2010). Horizontal transposon transfers (HTTs) have been detected in many eukaryotes, but the vast majority of HTTs were reported in Animalia and a small fraction, ~4%, were found in Plantae (Wallau et al. 2012). The horizontal movement of DNA transposons and LTR retrotransposons has been reported between species from different families and even phyla within the Animalia (Opisthokonta) and Plantae (Archaeplastida) kingdoms (Diao et al. 2006; Bartolome et al. 2009; Gilbert et al. 2010; Wallau et al. 2012; El Baidouri et al. 2014). However, much less is known about HTTs across multicellular eukaryotic supergroups or kingdoms and only one case of HTT between Animalia and Plantae has been reported (Lin et al. 2016). Here, we present the identification of a RTE non-LTR retrotransposons distributed across the green plant phylogeny and provide evidence for horizontal transfer of one RTE clade between arthropods and an ancestral angiosperm.

Results

Discovery and Structure of a New RTE Retrotransposon in a Wild Peanut

In the process of annotating TEs from the genome of the wild peanut species, *Arachis duranensis* (AA, $2n = 20$) (Bertioli et al. 2016), we identified a new non-LTR retrotransposon,

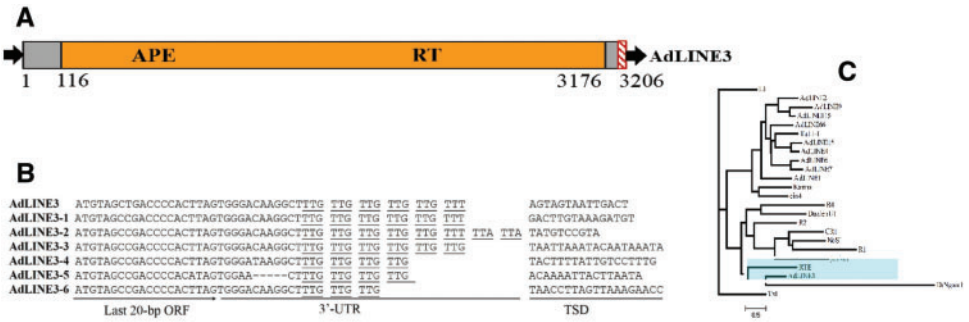

Fig. 1. Structural and phylogenetic analysis of AdLINE3 retrotransposon. (***A***) Schematic structure of AdLINE3. Gray boxes represent 5′ and 3′ untranslated regions (UTRs) and the orange indicates the coding region for the retrotransposase protein. Red stripes denote the 3′ terminal tandem repeats and the black arrows represent target site duplication (TSD). (***B***) Sequence alignment of the 3′ termini of AdLINE3 and other six complete members. The 3′ terminal tandem repeats are variable as shown by underlined triplets. The TSDs ranging from 10 to 19 bp were manually inspected and shown after the AdLINE3 sequences. (***C***) Phylogenetic tree of conserved reverse transcriptase (RT) from AdLINE3 and other LINEs. The RTE clade is marked by blue shading.

referred to as AdLINE3. This retrotransposon is 3,205 bp in size and has a single open reading frame (ORF) encoding a 1,019-amino acid polyprotein containing two functional domains for apurinic endonuclease (APE) and reverse transcriptase (RT). Unlike other complete long interspersed nuclear elements (LINEs) found in plants (Komatsu et al. 2003), the 3′ terminus of AdLINE3 lacks a poly-A tail and instead contains tandem repeats (fig. 1*A*) which were labile between different copies of the AdLINE3 family (fig. 1*B*). These structural features are similar to an RTE element in the nematode (*Caenorhabditis elegans*) (Malik and Eickbush 1998), but different from L1 and other clades of LINEs (Wicker et al. 2007). Sequence comparisons indicated that the AdLINE3 protein shares significant sequence similarity with the RTE retrotransposase protein (E value $= 8\times e^{-52}$). We further conducted phylogenetic analysis using the conserved RT domains from AdLINE3 and other non-LTR retrotransposons and found that AdLINE3 was grouped together with RTE element, whereas, other plant LINEs were grouped together (fig. 1*C*). Therefore, AdLINE3 represents a new member of the RTE clade.

Widespread Distribution and Recent Activity of RTE Retrotransposons in Plants

Thus far, nearly all RTE retrotransposons have been reported in animals including nematodes (Malik and Eickbush 1998) and vertebrates (Gilbert et al. 2010), and only a few RTE sequences were identified in plants (Zupunski et al. 2001; Mehra et al. 2015). We searched GenBank using AdLINE3 and an RTE retrotransposon from the nematode (Malik and Eickbush 1998), and identified highly similar sequences (E value $< 1\times e^{-10}$, >150 bp) in 81 plants ranging from angiosperms to algae (supplementary tables S1 and S2, Supplementary Material online). Notably, we identified complete RTE retrotransposons in 30 flowering plants, flanked by target site duplications (TSDs) of 8–20 bp and containing variable 3′-terminal motifs, but mostly TTG tandem repeats (supplementary table S1, Supplementary Material online). Thus, RTEs are found throughout the plant kingdom. Sequence comparisons of plant RTEs revealed that RTEs from other flowering plants shared significant sequence identity to AdLINE3 and among each other, however, they showed no significant sequence identity to algal RTEs. Thus, we hypothesize that the RTEs in flowering plant genomes were likely derived from a single ancestor that may be distinct from the ancestral algal RTEs, we refer hereafter to this family as An-RTE (Angiosperm RTE).

An-RTEs are abundant in many flowering plants indicative of massive amplifications (supplementary table S3, Supplementary Material online). We identified multiple complete An-RTEs that shared >98% sequence identity in apple (*Malus domestica*), soybean (*Glycine max*), stiff brome (*Brachypodium distachyon*), maize (*Zea mays*), and the two wild peanuts, *A. duranensis* and *A. ipaensis* (BB, $2n = 20$). In addition, we found numerous ESTs showing 98–99% sequence identity to the complete RTEs in all six species. As non-LTR retrotransposons move via a "copy and paste" model, high sequence identity may indicate recent activity. To gain insight into activity of the An-RTE retrotransposon, we investigated polymorphic insertions of AdLINE3 in the two wild peanut genomes, *A. duranensis* and *A. ipaensis*, that diverged from a common ancestor ~2.2 Ma (Bertioli et al. 2016). About 114 and 178 complete AdLINE3s defined by the tandem repeats at 3′ end and TSDs were detected in *A. duranensis* and *A. ipaensis*, respectively. About 32 and 26 new insertions were identified in *A. duranensis* and *A. ipaensis*, respectively, by comparing TSDs and flanking sequences. Other complete elements were either shared by the two genomes or inserted into repetitive regions.

Transposon display, which generates amplicons that target a specific transposon and flanking restriction sites (Casa et al. 2000), was used to detect retrotransposon polymorphisms. Polymorphic bands were detected between the two wild species (supplementary fig. S1*A* and *B*, Supplementary Material online) indicating that new insertions of AdLINE3 occurred in both species after their divergence. Nearly all bands generated in *A. ipaensis* were found in cultivated peanuts (*Arachis hypogaea*, AABB, $2n = 40$) which supports its suggested role as the B genome donor (Bertioli et al. 2016). Whereas many unique bands were detected in *A. duranensis* (V14167) suggestive that this accession is a close relative to the A-genome

accession that gave rise to the cultivated peanut (~9,400 years ago) (Bertioli et al. 2016) or that AdLINE3 was more active in *A. duranensis* than *in A. ipaensis*. In addition, polymorphic AdLINE3s were identified among seven cultivated peanut varieties (supplementary fig. S1*B*, Supplementary Material online) revealing more recent retrotranspositions of AdLINE3 in peanut. To detect transcriptional activation of AdLINE3, we conducted reverse transcription (RT)-PCR analysis with the primers targeting the reverse transcriptase of AdLINE3. A visible band was amplified in the stems, leaves, and flowers in both wild and cultivated peanuts, though signal was weaker in the leaves of *A. ipaensis* (supplementary fig. S1*C*, Supplementary Material online).

Identification of An-RTE Homologs in Animals

We searched the animal genomes deposited in GenBank with AdLINE3 and other An-RTEs, and identified homologous sequences (*E* value $< 1\times e^{-10}$) in 42 animals including one each from the Phyla Nematoda and Cidaroida, and 14 and 26 from Arthropoda and Chordata, respectively (supplementary table S4, Supplementary Material online). Among the identified sequences, we found only three complete An-RTE homologs, Ace-RTE2 in the zoonotic hookworm (*Ancylostoma ceylanicum*), San-RTE2 in the eyeless fish (*Sinocyclocheilus anshuiensis*) and Ban-RTE in the squinting bush brown butterfly (*Bicyclus anynana*). The majority of these homologs were fragmentary. Sequence comparisons indicate that homologs of AdLINE3 in both plants and animals shared >60% sequence identity across an over 250-bp region (supplementary fig. S2*A* and *B*, Supplementary Material online). However, RTEs from arthropods show higher sequence similarity to An-RTEs in plants over longer matching regions than those from fishes and nematodes, suggesting a closer relationship between An-RTEs and arthropod homologs than between An-RTEs and homologs in other animals.

To validate our computational analyses, we conducted PCR and sequence analyses for seven animals and ten flowering plants using primers targeting the flanking regions of the shared sequences (supplementary table S5, Supplementary Material online). The sequenced PCR products showed >99% identity to the identified RTE sequences in all organisms but mulberry (*Morus notabilis*) for which we detected 95% sequence identity between the PCR product and the RTE sequence from GenBank, likely due to sequencing another RTE copy or variation among accessions. We further conducted Southern blots using the amplified PCR products as probes. Strong signals were detected in silkworm (*Bombyx mori*) and corn earworm (*Helicoverpa zea*) but not in plants using the RTE sequences from insects as probes (fig. 2*A* and *B*). In addition, no hybridization signal was found in the animals when using plant RTE probes (fig. 2*C* and *D*). Thus, our PCR analyses and DNA hybridizations confirm the presence of An-RTE homologs in animals and exclude the possibility of plant DNA contamination. Plants and animals diverged from a common ancestor ~1,600 Ma (Meyerowitz 2002) and retrotransposons, particularly in plants, are highly dynamic sequences (Vitte et al. 2007; Gao et al. 2009). The significant

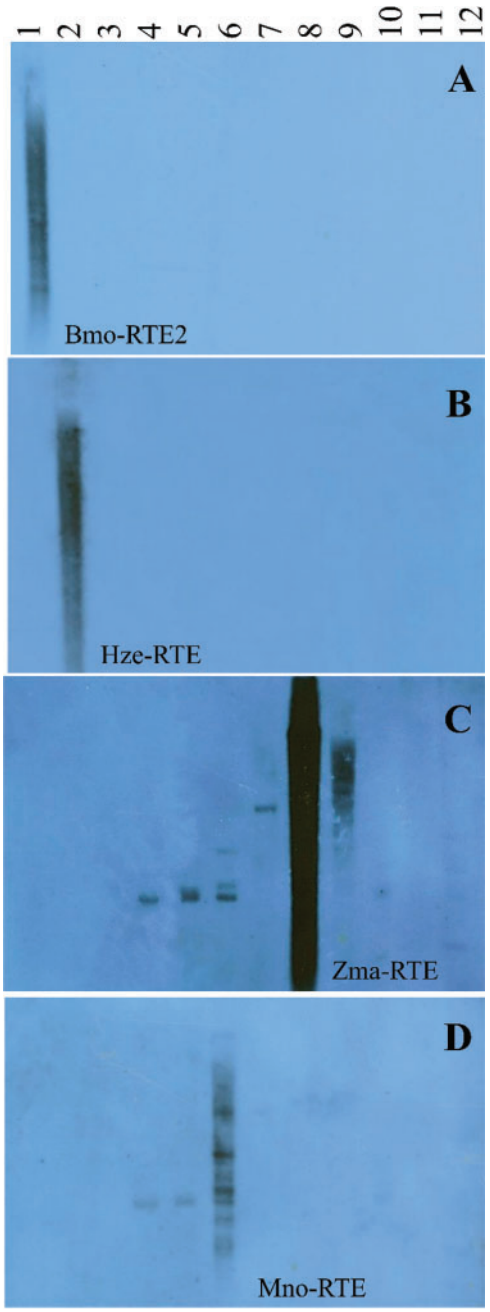

FIG. 2. Southern blot of genomic DNAs from nine plants and three animals. The RTE sequence from *Bombyx mori* (***A***), *Helicoverpa zea* (***B***), *Zea mays* (***C***), and *Morus notabilis* (***D***) were used as probes. Lanes 1–12 contain DNA from *Bombyx mori*, *Helicoverpa zea*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Glycine max*, *Gossypium hirsutum*, *Oryza sativa*, *Zea mays*, *Hordeum vulgare*, *Morus notabilis*, *Arachis hypogaea*, and *Brassica napus*, respectively.

sequence identity between retrotransposons from two different kingdoms strongly indicates potential horizontal transfer of RTEs between animals and plants.

Phylogenetic Analyses of RTE Retrotransposons in Animals and Plants

To gain insights into evolutionary relationships among RTEs in animals and plants, we identified RTEs in genomes from

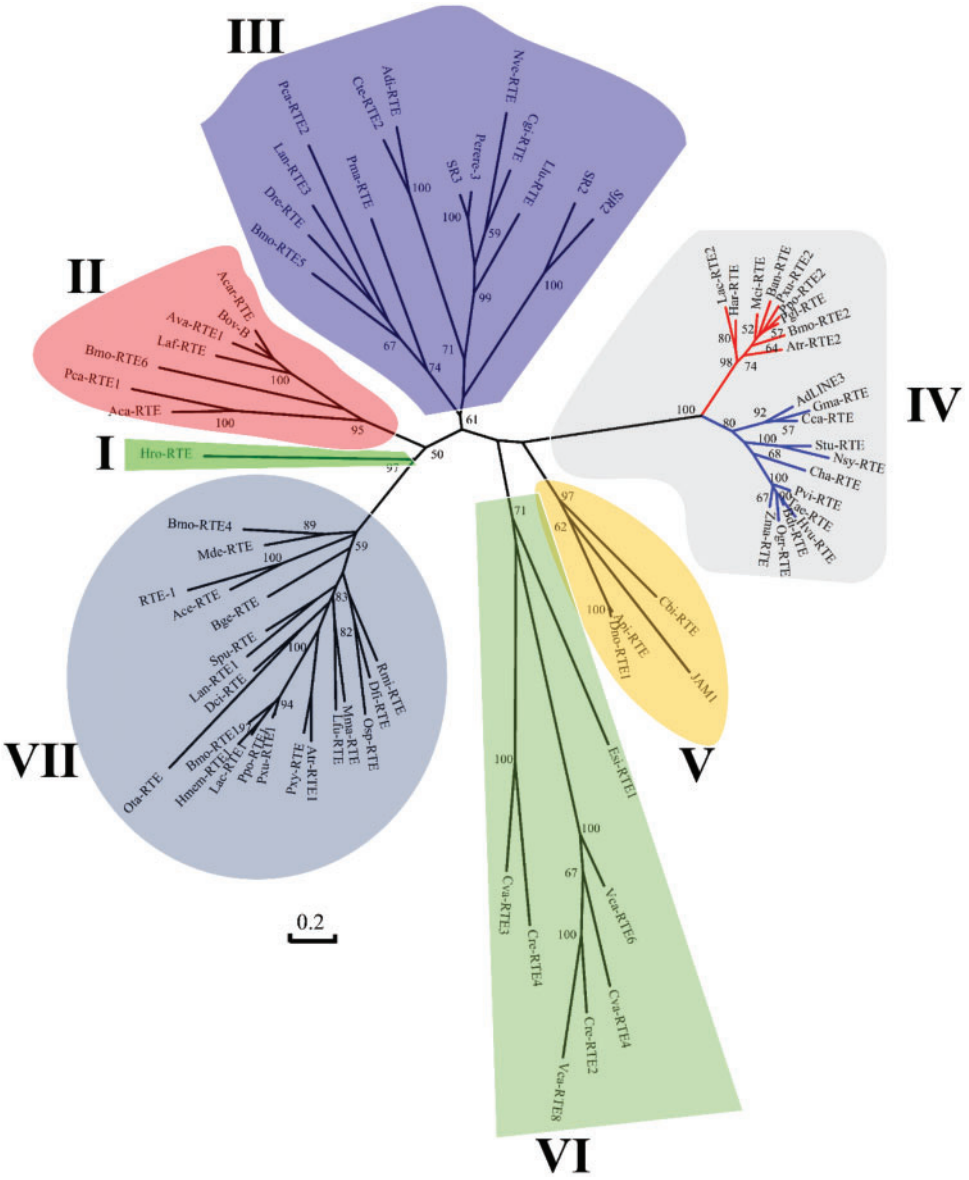

Fig. 3. Phylogenetic analysis of conserved RT domains. RAxML was used to generate an unrooted tree using 75 RTEs including 56 from animals, 12 from flowering plants, and 7 from algae. The bootstrap values of > 50% are shown. The blue and red branches in group IV indicate An-RTEs from flowering plants and the animal homologs, respectively.

sequenced animal phyla. A total of 95 RTEs including 42 exhibiting significant similarity to An-RTEs were identified from 12 phyla. In contrast to flowering plants where only one RTE family was identified, animal genomes often contain multiple RTE families, such as six families were identified in the silkworm genome (supplementary table S4, Supplementary Material online). In order to understand the diversity of RTE families and relationships among families, we aligned protein sequences of conserved RT domains using PASTA (Mirarab et al. 2015) and conducted phylogenetic analysis using RAxML (Stamatakis 2006). Our analysis included plant and animal RTEs identified in this study and seven previously reported animal RTEs (supplementary table S6, Supplementary Material online). The resulting phylogeny indicated that the RTEs from animals and plants were grouped into seven clades, all algae RTEs were grouped into a separate clade (clade VI). The An-RTEs from flowering plants were not grouped together with algal RTEs but were instead placed together with animal homologs in clade IV with 100% bootstrap values (fig. 3). The An-RTEs formed a subclade within clade IV sister to a subclade of their homologs identified in diverse animal genomes. The separation of algal RTEs and An-RTEs in the phylogeny raises the possibility of horizontal transfer of RTEs between flowering plants and some animals after the divergence of Archaeplastida and Opisthokonta lineages.

To further investigate into the evolutionary origin of the An-RTE family in flowering plants, we conducted sequence comparisons between An-RTEs and the animal homologs. Dno-RTE3 from Russian wheat aphid (*Diuraphis noxia*) and Cex-RTE from bark scorpion (*Centruroides exilicauda*) show lowest *E* values over longer matching regions against An-RTEs

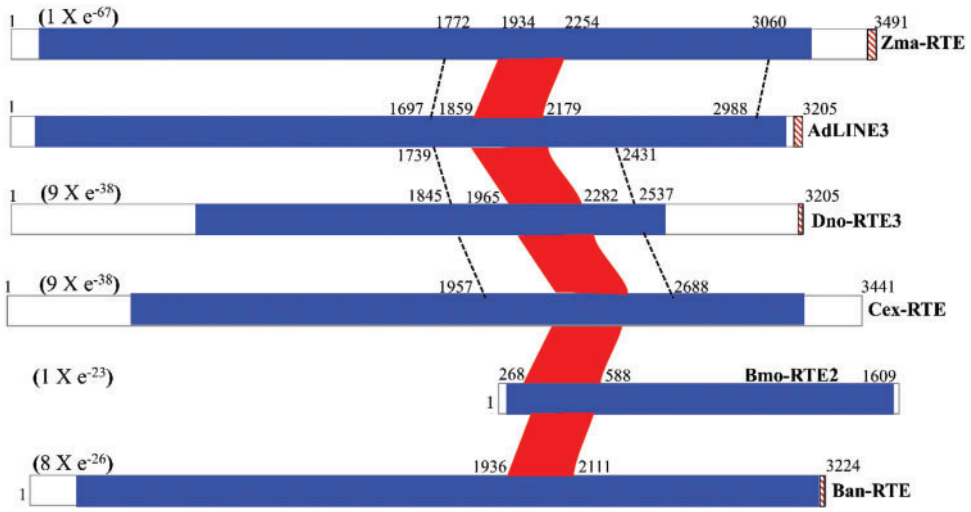

Fig. 4. Sequence alignment of AdLINE3 and homologs. Blue blocks show the retrotransposase-encoding regions. The conserved domains shared by AdLINE3 and all homologs in both animals and plants are shown in red. The Ban-RTE from the squinting bush brown butterfly and Bmo-RTE2 from silkworm share ~300 bp in the conserved region, but Zma-RTE from maize, Cex-RTE from bark scorpion and Dno-RTE3 from Russian wheat aphid share a longer region with AdLINE3, marked by black dotted lines. The numbers in () are the *E* values of a Blast2 search of AdLINE3 against Zma-RTE and RTE retrotransposons in animals (supplementary table S4, Supplementary Material online) identified here in plant and animals.

than other animal-derived RTEs (fig. 4). A phylogenetic analysis of ~300-bp conserved DNA sequences encoding a portion of the RTE reverse transcriptase (red region in fig. 4) indicated that An-RTEs formed a well-supported clade with homologous RTEs from aphids, Dno-RTE3, and Api-RTE2, and Cex-RTE from a bark scorpion (fig. 5). These sequences also exhibited a larger span of alignable sequence (fig. 4).

Genome-Wide Comparisons between Arthropods and Flowering Plants

Stochastic loss of ancestral sequences can also result in phylogenetic incongruence (Keeling and Palmer 2008). To test this possibility and provide additional evidence for horizontal transfer of RTEs, we conducted genome-wide comparisons of RTEs and genes between two plants, soybean and maize, a dicot and a monocot, respectively, and two arthropods, Russian wheat aphid and bark scorpion. If the RTE family was vertically transmitted and maintained by neutral evolutionary process in animal and plant genomes, it must have been present in the ancestor of arthropods and plants. Thus, the number of synonymous substitution rate (Ks) of the RTEs should be equal or greater than that of vertically transmitted homologous gene sequences (HGSs) (Pace et al. 2008; Wallau et al. 2012), If as expected, TEs evolve more quickly than genes (Vitte et al. 2007; Gao et al. 2009), the sequence identity of RTEs between animals and plants should be lower than that of HGSs. All annotated genes were used to identify 3,400 pairs of homologous genes between the four plant and arthropod genomes. However, the vast majority (~90%) of homologous genes showed no significant DNA sequence similarity, and the average nucleotide sequence identity of all homologous genes between the arthropods and plants ranged from 5.9% to 8.8%, much lower than the genome-wide comparisons between the An-RTEs from soybean and maize and Dno-RTE3 in Russian wheat aphid and Cex-RTE in bark scorpion that ranged from 53.8% to 68.2% (fig. 6A).

We next investigated the Ks values for pairwise comparisons of genes and RTEs between the plants and arthropods. The Ks values of genes from all four plant and arthropod combinations show nearly normal distributions with mean values of 1.64, 1.71, 1.83, and 1.90 between bark scorpion/soybean, bark scorpion/maize, wheat aphid/soybean, and wheat aphid/maize, respectively. In contrast to genes, the Ks values of RTEs are not typically normally distributed, and the mean values were 1.12, 1.17, 0.91, and 1.11 between Cex-RTE/Gma-RTE, Cex-RTE/Zma-RTE, Dno-RTE3/Gma-RTE, and Dno-RTE3/Zma-RTE, respectively (fig. 6B). We conducted Wilcoxon test and found that the distribution of Ks values of homologous RTEs was significantly different from genes ($P < 2.2e^{-16}$) in all plant–arthropod comparisons. The comparisons of sequence identity and synonymous divergence rates revealed that An-RTEs and their arthropod homologs had lower sequence divergence values than that of HGSs. This again supports the proposition that the RTE family was transmitted horizontally between arthropods and flowering plants.

Contribution to Gene Structures

Transposons transferred between distantly related organisms are difficult to maintain as they likely undergo selective pressure to be removed from the recipient genomes over time as they may be harmful to the host. The exception to this would be if they provide some selective advantage to either themselves or the recipients, such as contributing to a biochemical network (Boto 2010; Soucy et al. 2015). An-RTEs have been retained in plant genomes for long time. To provide insights into the long maintenance of An-RTEs, we searched the coding DNA sequences (CDSs) of maize and soybean with Zma-RTE and Gma-RTE, two maize genes, and four soybean genes show significant sequence identity to the

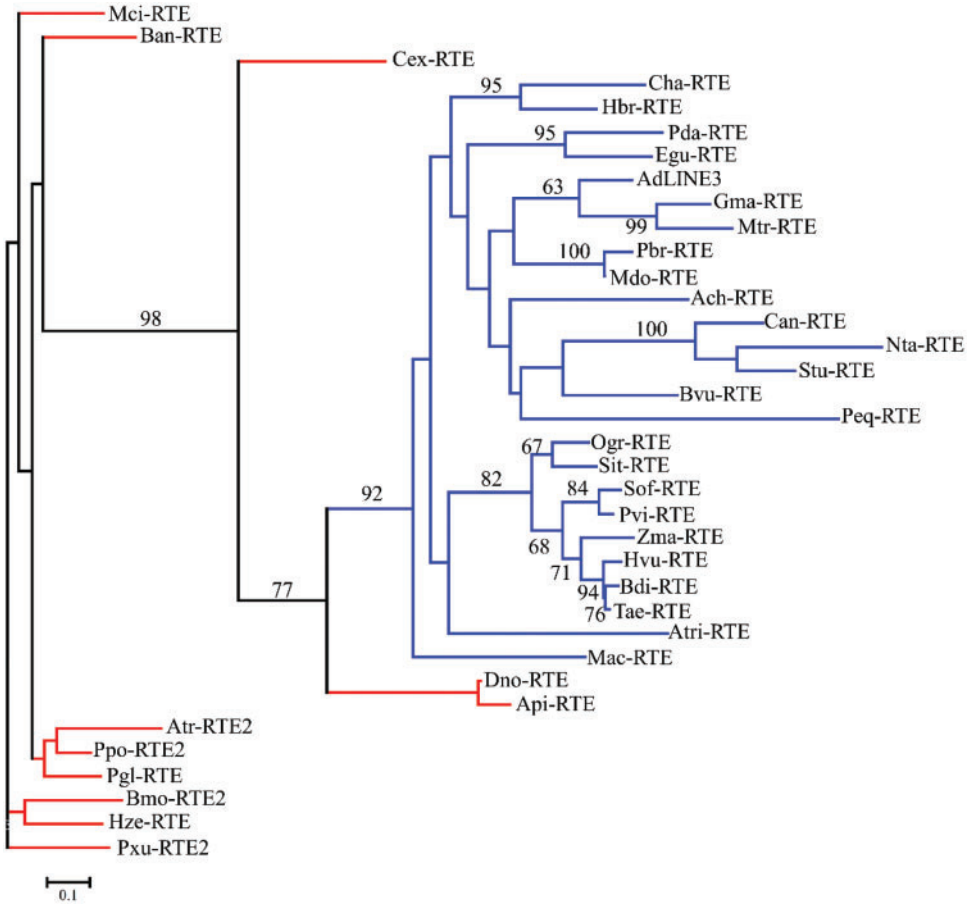

Fig. 5. Phylogenetic analysis of conserved DNA sequences of plant and animal RTEs. The RAxML tree was built from PASTA alignment of ~300-bp conserved sequences from 25 An-RTEs from flowering plants (blue) and 11 homologs from animals (red). The bootstrap values of > 50% are shown.

An-RTEs (*E* value $< 1\times e^{-9}$). All these genes encode enzymes catalyzing or recognizing biochemical products (supplementary table S7, Supplementary Material online). For example, soybean gene LOC100797314 contains a 584-bp Gma-RTE sequence spanning the third and fourth introns and fourth exon (supplementary fig. S3, Supplementary Material online), and encodes an LRR receptor-like serine/threonine-protein kinase that can interact with a diverse group of proteins and promote pathogen recognition (Afzal et al. 2008). These results demonstrate that the vast majority of An-RTEs in maize and soybean are located in noncoding regions, including intronic sequences, but a few An-RTEs serve as coding sequences for metabolic genes. We next compared the six genes with their homologs and estimated the ratios of nonsynonymous to synonymous substitution per site (Ka/Ks) The Ka/Ks values for all genes were less than one (supplementary table S7, Supplementary Material online) indicating that these genes have undergone purifying selection.

Discussion

Thus far, nearly all HTTs found in both plants and animals were related to DNA transposons and LTR retrotransposons (Diao et al. 2006; Bartolome et al. 2009; Gilbert et al. 2010; Wallau et al. 2012; El Baidouri et al. 2014) with only reported instances of non-LTR retrotransposons in animals (Walsh et al. 2013). This difference was due likely to the retrotransposition mechanism by which non-LTR retrotransposons including LINEs and SINEs nick host chromosomes and integrate the single-stranded RNA transcript onto the target sites that is more degradable and less stable than the double-stranded DNA intermediate used for movements of DNA transposons and LTR retroelements (Schaack et al. 2010; Wallau et al. 2012).

We hypothesize that an ancestral of all flowering plants acquired an RTE retrotransposon from arthropods and not the converse. This hypothesis is supported by the following observations: 1) An-RTEs were more closely related to their arthropod homologs than algal RTEs (figs. 3 and 5); 2) multiple and diverse RTE families were found in arthropods and other animals indicating a long history in animals, but only a single RTE family was identified in flowering plants; and 3) An-RTE-like RTEs sampled from arthropods share a broad region of significant sequence similarity with An-RTEs (fig. 4).

To our knowledge, this represents the first evidence of potential horizontal transfer of non-LTR retrotransposon across animals and flowering plants, thus the exchange of single-stranded RNA can occur between organisms that diverged ~1,600 Ma (Meyerowitz 2002). A previous study revealed that the common ancestor of conifers obtained a

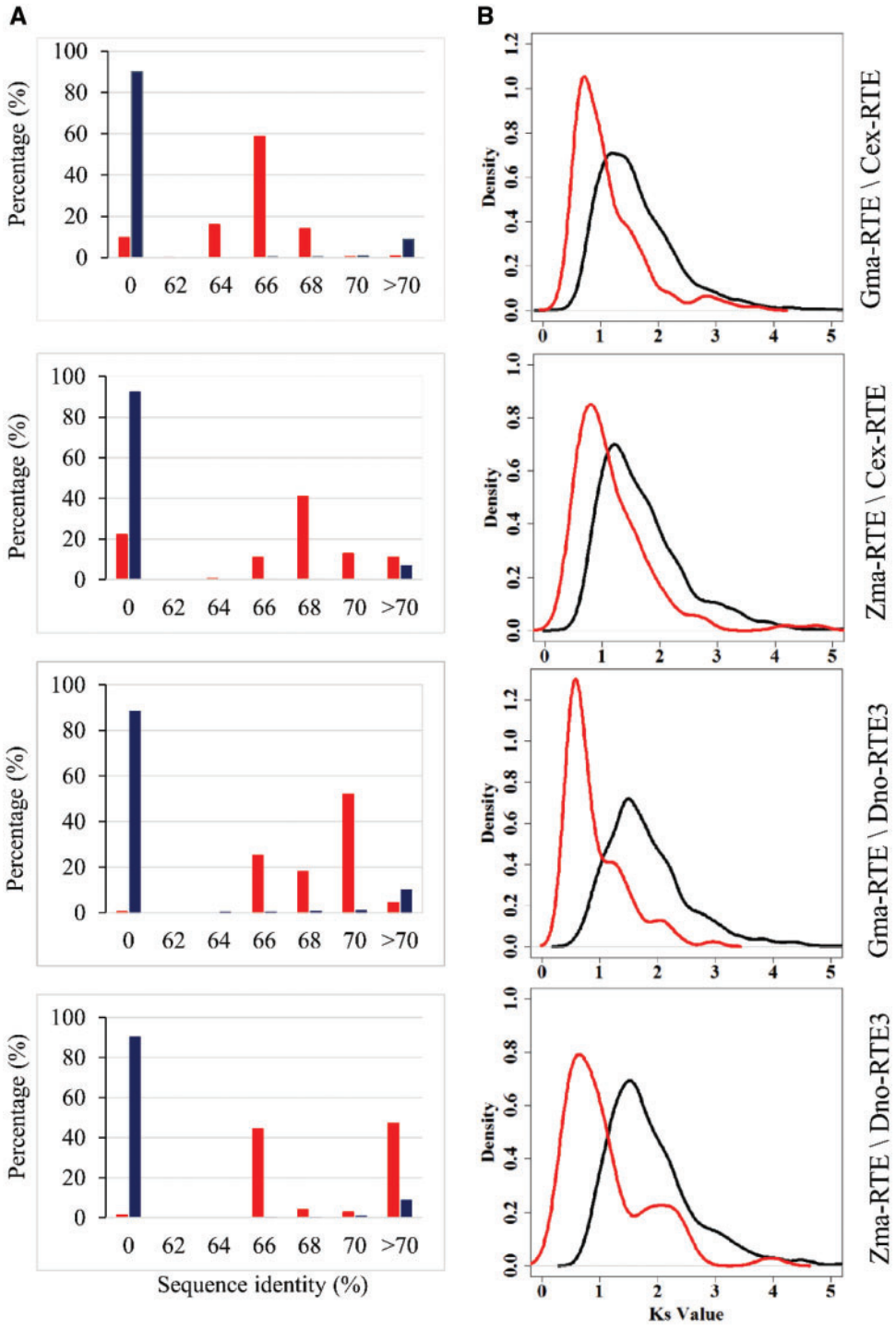

Fig. 6. Sequence comparisons of RTE retrotransposons and genes between two flowering plants (*Zea mays* and *Glycine max*) and two arthropods (*Diuraphis noxia* and *Centruroides exilicaud*). (***A***) The distribution of sequence identities of RTEs (red) and genes (blue) and (***B***) The distribution of Ks values of RTEs (red) and genes (blue).

penelope-like retroelement from arthropods ~340 Ma (Lin et al. 2016), the only reported case of HTT across Animalia and Plantae thus far. Interestingly, this transfer from an arthropod corresponds with our observation that arthropods were the likely donor for the RTEs reported here. The transfer of the penelope retroelement occurred early during insect evolution (~340 Ma) (Lin et al. 2016). Our phylogenetic analysis implies that An-RTE-like elements in arthropod genomes were transferred to an ancestor of flowering plants (fig. 5) after divergence from gymnosperms. No RTEs were detected in available land plant genomes outside of the angiosperms including *Physcomitrella patens*, *Selaginella moellendorffii*, and Norway spruce. However, this hypothesis must be rigorously tested as more nonangiosperm plant genomes are sequenced.

Sequencing of more arthropod genomes will also inform our understanding of HTT between plants and associated arthropod species. It is intriguing that An-RTEs are most closely related to aphid RTEs. Aphids are phloem-sucking insects that have been coevolving with flowering plants for at least 150 Ma (Peccoud et al. 2010). Aphids are known to transmit viruses to their hosts (Tamborindeguy et al. 2010). It is possible that viruses or other microbes such as bacteria and fungi carried by aphids served as the vector or intermediate hosts for the movement of RTE retrotransposon to flowering plants. Virus-mediated HTT between insects and microbes have been described previously (Dunning Hotopp et al. 2007; Gilbert et al. 2014). We also cannot rule out the possibility that other arthropod(s) not yet sequenced or no longer extant may have served as the donor.

Transposons are the most abundant sequences in flowering plants and have played crucial roles in genomic novelty and variation as they can change genome sizes and structures, contribute to the creation of new genes and gene regulatory networks (Feschotte 2008; Bennetzen and Wang 2014). However, the origin of transposons in plant genomes is still unclear. For example, LINEs were identified in numerous plants and shared phylogenetic relationship with L1 retrotransposons in mammalians (Komatsu et al. 2003), however, little is known about how and when L1-like retrotransposons emerged in plants. The evolutionary origin of An-RTEs seems to be distinct from other types of transposons in plants such as miniature inverted-repeat transposable elements (MITEs) and terminal-repeat retrotransposons in miniature (TRIMs) that were generated by internal deletions of large endogenous transposons (Gao et al. 2016). Numerous HGTs were found in prokaryotes, but only a few HGTs were identified between the kingdoms Animalia and Plantae, including a controversial case in which putatively transferred genes cannot be found in germ cells (Bhattacharya et al. 2013). The low rate of HGT between herbivorous animals and plants is likely due to reduced efficiency of homologous recombination between divergent organisms (Soucy et al. 2015), HGTs between organisms separated by long evolutionary distances are constrained by surveillance systems and highly divergent regulatory networks in the recipient (Boto 2010). Analysis of available genome data suggest that ancestral An-RTE elements were transferred to an ancestral angiosperm genome and subsequently maintained and allowed to proliferate in flowering plant genomes. The mechanism behind the unusual sequence conservation of An-RTEs within angiosperm genomes is still not clear. The copy numbers and concentration in noncoding regions suggest that these elements have maintained amplification activity for some time and that their insertions may have had little deleterious effects or were neutral for the host. This provides a novel perspective on the emergence, maintenance, and domestication of new transposon in the genomes of host plants.

Arthropods are important as they provide foods for human and other animals, serve as pollinators for plants and play other ubiquitous roles in ecosystems such as decomposers by feeding on dead animals or other waste (Sander and van Veen et al. 2011). However, some arthropods are harmful as they feed on plants and can transmit disease-causing viruses or other microbes to plants. The fossil records indicated that insects have been feeding on plants for >410 My (Labandeira and Currano 2013) and point to a long history of coevolutionary relationships between plants and herbivorous or pollinating insects. However, we know little about the exchange of genetic material between plants and arthropods during this coevolutionary period. Previously studies have shown that some arthropods such as *Rhodnius prolixus* and *Amblyomma* likely served as vectors for widespread HTTs between vertebrates (Gilbert et al. 2010; Walsh et al. 2013). Our results together with the previous analysis of penelope retroelements in conifer genomes (Lin et al. 2016) suggest that plant genomes have also acquired new transposons from arthropods. Acquisitions of foreign genes can provide the recipient some fitness benefits including disease resistance or other adaption traits (Zhu and Gao 2014). The An-RTEs have persisted in flowering plant genomes for extraordinarily long time periods, and some An-RTE-derived sequences have been recruited to function in enzyme-coding genes, suggesting that these foreign elements may play roles in plant genome evolution and gene function.

In conclusion, our results indicate widespread distributions of RTE retrotransposons in the plant kingdom and provide evidence that the An-RTE family in flowering plants were acquired from arthropods via ancient horizontal transposon transfer. Our data also suggest arthropods were the contributors of foreign genetic material for plant genomes and insect-mediated HTTs has impacted plant gene/genome structure and evolution.

Materials and Methods

Plant and Animal Materials

All plants used in this study were collected in the greenhouse or experimental fields at the University of Georgia except for mulberry leaves obtained from the Shandong Academy of Agricultural Sciences. The larvae or adult insects of *Bicyclus anynana*, *Helicoverpa zea*, and *Drosophila melanogaster* were obtained from Drs. Antónia Monteiro, Dawn Olson, and Cordula Schulz, respectively.

Identification and Copy Number Estimation of Non-LTR Retrotransposons

To annotate non-LTR retrotransposons in peanut genomes, the proteins from L1, RTE-1, and other superfamilies of LINEs (supplementary table S6, Supplementary Material online) were used as queries to search against Moleculo-derived long reads from the wild peanut species *A. duranensis* (3.2 Gb, 773,616 reads varying in size from 1,500 to 22,045 bp with the average size of 4,121 bp) and *A. ipaensis* (8.1 Gb, 2,004,936 reads ranging from 1,500 to 19,943 bp with the average size of 4,054 bp) (Bertioli et al. 2016). All significant sequences (E value $< 10^{-5}$) were extracted and manually inspected and complete LINEs defined on TSD and terminal motifs. To identify RTE retrotransposons in other plants and animals, the DNA, and protein sequences of AdLINE3 and the RTE retrotransposon from *Caenorhabditis*

elegans (Malik and Eickbush 1998) were used as queries to conduct BLASTN and TBLASTN searches against database in GenBank. The significant hits and their flanking regions (5 Kb for each side) were extracted and inspected to examine boundaries and TSDs. We excluded hits <150 bp. To estimate the copy number of retrotransposons, the identified RTEs were used to screen their host genomes with Repeatmasker (http://www.repeatmasker.org). The program was run using the default parameters but "nolow" option. The transposon copy numbers were summarized with a custom script, and overlapping regions in the RepeatMasker output file were counted only once.

PCR, RT-PCR and Sequencing

PCR and RT-PCR amplifications were performed following our previous protocol (Gao et al. 2016). Briefly, 20 ng of genomic DNA was used to amplify the targeted sequences in 25 μl reactions, 5 μl PCR reactions were taken to check the amplification, and the remaining products were purified with the Qiaquick PCR purification kit (QIAGEN, Venlo, Netherlands) or cloned with the TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced by the Sanger method. Four micrograms total RNA from each sample was converted into single-strand cDNA with reverse transcriptase (Invitrogen, Carlsbad, CA). The cDNA reactions were then diluted 4- to 5-fold, and 2 μl of the diluted cDNA were used as templates for PCR amplifications with the primers targeted to the reverse transcriptase of AdLINE3 and actin gene in peanut. The primers for PCR and RT-PCR are listed in the supplementary table S5, Supplementary Material online.

Southern Blot

To perform DNA hybridization, 5 μg genomic DNA for each sample was digested with EcoR I (New England, Ipswich, MA) at 37 °C for 8 h. The digested DNAs were separated by electrophoresis on a 0.8% (w/v) agarose gel at 45 v for overnight and transferred onto a Hybond N+ membrane (GE Healthcare, Little Chalfont, UK). Primers targeting the reverse transcriptase region of the RTEs (supplementary table S5, Supplementary Material online) were used to amplify the fragments. About 500-ng purified PCR products were used as probes and labelled with the DIG DNA labeling kit (Roche, Mannheim, Germany). Hybridization was performed using the "DIG easy hyb" system by following the manufactures instructions.

Transposon Display

Transposon display was conducted as described (Casa et al. 2000). In brief, 150-ng genomic DNA from each of ten peanut accessions was digested with *MseI* at 37 °C for 2 h, and the digested DNAs was mixed with the adapters at 4 °C overnight. Preselective amplification was performed using the adaptor primer (5′-GATGAGTCCTGAGTAA-3′) and the AdLINE3 specific primer (5′-GAAGACCTAAGAAGACCATC-3′). PCR reaction for selective amplification was conducted using the adapter primer and AdLINE3 primer (5′-GAAGACCTAAGAAGACCATC-3′, where N was A, T, C, or G). A 6.5% polyacrylamide gel was prerun at 1500 V for 20 min on a *LI-COR* 4300 DNA Analyzer. Samples (0.5 µl each) were loaded onto a gel and then run for 3.5 h at 1500 V. Image was viewed with both the 700 and 800 channels.

Phylogenetic Analysis of Retrotransposons

The proteins of all non-LTR retrotransposons identified in this study and the published LINEs (supplementary table S6, Supplementary Material online) were analyzed with Fgenesh gene-finder (http://linux1.softberry.com) and GENSCAN (http://genes.mit.edu/GENSCAN.html). The annotated proteins were used to conduct BLASTP searches to determine the conserved RT domains. The conserved DNA domains of An-RTEs and their homologs in animals were determined based on sequence alignments of the RTEs. To build the phylogenetic trees, nucleotide, and protein sequences were aligned using PASTA (Mirarab et al. 2015) and phylogenetic analyses were performed using RAxML (Stamatakis 2006), with 200 bootstrap replicates and either GTRGAMMA or PROTGAMMAWAGF models of substitution for nucleotide and protein sequences, respectively.

Homologous Sequence Analysis

We downloaded annotated genes and genome sequences from NCBI (http://www.ncbi.nlm.nih.gov), Phytozome (https://phytozome.jgi.doe.gov), and other websites (supplementary table S8, Supplementary Material online). To identify homologous genes, the proteins of all annotated genes in each genome were used to conduct BLASTP searches (*E* value $< 10^{-5}$) against other genomes. The best subject alignment (lowest *E* value) for each query sequence and the best query alignment for each subject sequence were then compared. The best query and subject sequences for a given pair of sequences were considered homologous genes. The DNA sequences corresponding to the homologous gene pairs were then aligned using BLASTN with the default parameters except for an *E* value of $<1\times10^{-5}$ and reward for a nucleotide match of 2. The gene sequence identities were summarized based on the distribution of best hit (lowest *E* value) sequence identity for each pair. Sequence identity was considered 0 if a homologous gene pair showed no significant sequence similarity and/or the matched regions were <50 bp.

To estimate the sequence identity between An-RTEs in soybean and maize, and their homologs in Russian wheat aphid and bark scorpion, we extracted all "complete" An-RTE sequences in soybean and maize containing all retrotransposase-encoding domains but we allowed deletions <50-bp at either the 5′ end or 3′ end. Complete RTEs were not found in the two arthropods, thus, only sequences that covered >75% of the reference RTEs were used to conduct BLASTN analyses against the extracted plant An-RTEs. The parameters of BLASTN and data summary of RTE retrotransposons were the same as for genes.

Calculations of Ks Values of Homologous Genes and RTE Retrotransposons

The proteins of all annotated genes in each species were used to search against other genomes using INPARANOID 4.1 that uses the pairwise similarity scores, calculated using

NCBI-Blast, between two complete proteomes for constructing homology/orthology groups (Remm et al. 2001), and only gene pairs with bootstrap value of 100% were retained. For each gene pair, the CDSs from both species were aligned by "Clustalw" and Ks values were calculated using the "Bio:: Align:: DNAStatistics" BioPerl module (Stajich et al. 2002). All extracted "complete" An-RTE sequences from soybean and maize and the extracted RTE sequences in the wheat aphid and bark scorpion were used to calculate Ks values by comparisons of the annotated CDSs using the same Perl script for genes. Statistical analysis and density distributions were performed using R Project for Statistical Computing (https://www.r-project.org).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Drs Antónia Monteiro (National University of Singapore), Dawn Olson (USDA), Gary Puterka (USDA), Stephen Richards (Baylor College of Medicine), Peng Chee (University of Georgia), Soraya Bertioli (University of Georgia), and other scientists for providing the tissues, DNAs and RNAs of animals and plants. This work was supported by the National Science Foundation (MCB 1339194) to S.A.J. and USA Peanut Foundation (047730-01) to S.A.J. and D.Y.G.

Author Contributions

D.Y.G. and S.A.J. designed the experiment. D.Y.G., C.M.X., and B.A. performed computational analysis. H.X. performed sequencing and DNA hybridization. Y.C. and P.O.-A. conducted TD analysis. K.H. and J.L.-M. built conducted phylogenetic analyses with PASTA and RAxML. D.Y.G. and S.A.J. wrote the article.

References

Afzal AJ, Wood AJ, Lightfoot DA. 2008. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Mol Plant Microbe Interact*. 21(5):507–517.

Bartolome C, Bello X, Maside X. 2009. Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. *Genome Biol*. 10(2):R22.

Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol*. 65:505–530.

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, et al. 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet*. 48(4):438–446.

Bhattacharya D, Pelletreau KN, Price DC, Sarver KE, Rumpho ME. 2013. Genome analysis of elysia chlorotica egg DNA provides no evidence for horizontal gene transfer into the germ line of this kleptoplastic mollusc. *Mol Biol Evol*. 30(8):1843–1852.

Boto L. 2010. Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci*. 277(1683):819–827.

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR. 2000. The MITE family heartbreaker (Hbr): molecular markers in maize. *Proc Natl Acad Sci U S A*. 97(18):10083–10089.

Dagan T, Artzy-Randrup Y, Martin W. 2008. Modular networks and cumulative impact of lateral gene transfer in prokaryote genome evolution. *Proc Natl Acad Sci U S A*. 105(29): 10039–10044.

Diao X, Freeling M, Lisch D. 2006. Horizontal transfer of a plant transposon. *PLoS Biol*. 4(1):e5.

Dunning Hotopp JC, Clark ME, Oliveira DCSG, Foster JM, Fischer P, Muñoz Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, et al. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317(5845): 1753–1756.

El Baidouri M, Carpentier M-C, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O, et al. 2014. Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res*. 24(5):831–838.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 9(5):397–405.

Gao D, Gill N, Kim H-R, Walling JG, Zhang W, Fan C, Yu Y, Ma J, SanMiguel P, Jiang N, et al. 2009. A lineage-specific centromere retrotransposon in *Oryza brachyantha*. *Plant J*. 60(5): 820–831.

Gao D, Li Y, Kim KD, Abernathy B, Jackson SA. 2016. Landscape and evolutionary dynamics of terminal repeat retrotransposons in miniature in plant genomes. *Genome Biol*. 17:7.

Gilbert C, Schaack S, Pace JK 2nd, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464(7293):1347–1350.

Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A, Herniou EA, Cordaux R. 2014. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat Commun*. 5:3348.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet*. 9(8):605–618.

Komatsu M, Shimamoto K, Kyozuka J. 2003. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. *Plant Cell* 15(8):1934–1944.

Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol*. 55:709–742.

Labandeira CC, Currano ED. 2013. The fossil record of plant-insect dynamics. *Annu Rev Earth Planetary Sci*. 41(1):287–311.

Lin X, Faridi N, Casola C. 2016. An ancient transkingdom horizontal transfer of penelope-like retroelementsfrom arthropods to conifers. *Genome Biol Evol*. 8(4):1252–1266.

Malik HS, Eickbush TH. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Mol Biol Evol*. 15(9):1123–1134.

Mehra M, Gangwar I, Shankar R. 2015. A deluge of complex repeats: the solanum genome. *PLoS One* 10(8):e0133962.

Meyerowitz EM. 2002. Plants compared to animals: the broadest comparative study of development. *Science* 295(5559):1482–1485.

Mirarab S, Nguyen N, Guo S, Wang LS, Kim J, Warnow T. 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J Comput Biol*. 22(5):377–386.

Pace JK 2nd, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon inmammals and other tetrapods. *Proc Natl Acad Sci U S A*. 105(44):17023–17028.

Peccoud J, Simon J-C, von Dohlen C, Coeur d'acier A, Plantegenest M, Vanlerberghe-Masutti F, Jousselin E. 2010. Evolutionary history of aphid-plant associations and their role in aphid diversification. *C R Biol*. 333(6–7):474–487.

Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol*. 314(5):1041–1052.

Sanders D, van Veen FJ. 2011. Ecosystem engineering and predation: the multi-trophic impact of two ant species. *J Anim Ecol*. 80(3):569–576.

Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol*. 25(9):537–546.

Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat Rev Genet*. 16(8):472–482.

Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JGR, Korf I, LappH, et al. 2002. The bioperl toolkit: Perl modules for the life sciences. *Genome Res*. 12(10):1611–1618.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.

Tamborindeguy C, Monsion B, Brault V, Hunnicutt L, Ju HJ, Nakabachi A, Van Fleet E. 2010. A genomic analysis of transcytosis in the pea aphid, *Acyrthosiphon pisum*, a mechanism involved in virus transmission. *Insect Mol Biol*. 19:259–272.

Vitte C, Panaud O, Quesneville H. 2007. LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomics* 8:218.

Wallau GL, Ortiz MF, Loreto ELS. 2012. Horizontal transposon transfer in Eukarya: detection, bias, and perspectives. *Genome Biol Evol*. 4(8):801–811.

Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL. 2013. Widespread horizontal transfer of retrotransposons. *Proc Natl Acad Sci U S A*. 110(3):1012–1016.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 8(12):973–982.

Zhu S, Gao B. 2014. Nematode-derived drosomycin-type antifungal peptides provide evidence for plant-to-ecdysozoan horizontal transfer of a disease resistance gene. *Nat Commun*. 5:3154.

Zupunski V, Gubensek F, Kordis D. 2001. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol Biol Evol*. 18(10):1849–1863.