LETTERS

A role for host–parasite interactions in the horizontal transfer of transposons across phyla

Clément Gilbert[1]*, Sarah Schaack[1]*, John K. Pace II[1], Paul J. Brindley[2] & Cédric Feschotte[1]

Horizontal transfer (HT), or the passage of genetic material between non-mating species, is increasingly recognized as an important force in the evolution of eukaryotic genomes[1,2]. Transposons, with their inherent ability to mobilize and amplify within genomes, may be especially prone to HT[3–7]. However, the means by which transposons can spread across widely diverged species remain elusive. Here we present evidence that host–parasite interactions have promoted the HT of four transposon families between invertebrates and vertebrates. We found that *Rhodnius prolixus*, a triatomine bug feeding on the blood of various tetrapods and vector of Chagas' disease in humans, carries in its genome four distinct transposon families that also invaded the genomes of a diverse, but overlapping, set of tetrapods. The bug transposons are ~98% identical and cluster phylogenetically with those of the opossum and squirrel monkey, two of its preferred mammalian hosts in South America. We also identified one of these transposon families in the pond snail *Lymnaea stagnalis*, a cosmopolitan vector of trematodes infecting diverse vertebrates, whose ancestral sequence is nearly identical and clusters with those found in Old World mammals. Together these data provide evidence for a previously hypothesized role of host–parasite interactions in facilitating HT among animals[3,7]. Furthermore, the large amount of DNA generated by the amplification of the horizontally transferred transposons supports the idea that the exchange of genetic material between hosts and parasites influences their genomic evolution.

In order to examine the factors underlying HT among widely diverged taxa, we began our investigation with *SPACE INVADERS* (*SPIN*), a recently described DNA transposon that has undergone repeated episodes of HT across the genomes of seven tetrapod lineages[5]. We first performed a series of BLASTN searches using the *SPIN* superconsensus sequence[5] as a query against all GenBank databases (Methods), including 102 species for which whole genome shotgun sequences are available. In addition to the vertebrates previously known to harbour *SPIN*, we found highly significant hits (*e*-values as low as 0, corresponding here to 86% identity over >1 kilobase) in the triatomine bug, *R. prolixus*, an hemipteran insect that feeds on the blood of mammals, birds and reptiles and serves as a vector for trypanosomes, the causal agent of Chagas' disease[8]. Significant hits were also obtained from multiple expressed sequence tag sequences generated for the freshwater snail *L. stagnalis*, which is an intermediate host for numerous trematodes parasitizing diverse vertebrate species[9].

The discovery of *SPIN* in two invertebrates associated with parasitic lifecycles was intriguing, especially because triatomines are known to feed on several species in which *SPIN* was previously identified[5]. Thus, we expanded our investigation to look for evidence of HT of additional DNA transposons between *R. prolixus* and vertebrates by performing BLASTN searches against the whole genome shotgun sequence of *R. prolixus* using a comprehensive collection of DNA transposons previously identified in vertebrates (Methods). These searches yielded significant hits (*e*-values range from 0 to 3×10^{-108}) for three families of mammalian DNA transposons: *hAT1*[10], *OposCharlie1* (*OC1*)[11] and *ExtraTerrestrial* (*ET*; see Methods for details on nomenclature). We confirmed the presence of all four transposon families in *R. prolixus* and of *SPIN* in *L. stagnalis* by PCR (polymerase chain reaction) amplification from genomic DNA and sequencing of cloned PCR products (Supplementary Table 1). We constructed consensus (that is, ancestral) sequences for each family of elements (*SPIN*, *OC1*, *hAT1* and *ET*) in every species, based on a multiple alignment of copies extracted from the database. Phylogenetic analysis of consensus transposase sequences shows that like *SPIN*, the other families identified (*OC1*, *hAT1* and *ET*) belong to the *hAT* superfamily, but are only distantly related to each other (Supplementary Fig. 1).

In the case of *SPIN*, an alignment of the *L. stagnalis* and *R. prolixus* consensus sequences with those generated previously for seven vertebrates[5] revealed an extremely high level of identity between invertebrates and vertebrates across the entire length of the consensus sequences (up to 98.4% between *L. stagnalis* and bat and up to 95.3% between *R. prolixus* and opossum; Supplementary Table 2; see also Supplementary Table 3). Phylogenetic analyses of multiple individual copies extracted from each species revealed unrooted trees with a star topology and no subfamily structure (Supplementary Fig. 2). This is indicative of the accumulation of discrete substitutions in each copy and consistent with neutral evolution of transposons after their integration in the genome[12] (see Methods for additional data supporting neutral evolution). Thus, the level of sequence conservation between the invertebrate and vertebrate *SPIN* sequences is incompatible with vertical inheritance from their common ancestor, which occurred >500 million years (Myr) ago. Instead, these data indicate that *SPIN* was able to invade the two invertebrate lineages horizontally and independently, as it did in each of the seven vertebrate lineages[5].

As for *SPIN*, several lines of evidence suggest that the taxonomic distribution of *hAT1*, *OC1* and *ET* is the result of independent HT of these elements in each of the vertebrate lineages where they were identified (Fig. 1). These include: (1) very high levels of sequence identity among species (ranging from 83.6% to 98.9% for all pairwise comparisons between consensus sequences; Supplementary Table 2; see also Supplementary Table 3 for comparison of individual copies), (2) no evidence for orthologous insertions among species (Supplementary Fig. 3), and (3) an inferred timing of transposon amplification clearly postdating the divergence of any two species harbouring these transposons (Fig. 1). Remarkably, *OC1* was able to infiltrate independently the germline of (at least) two prosimians and two anthropoid species, making it, to our knowledge, the most promiscuous and youngest DNA transposon ever identified in primates (~18 Myr ago in squirrel monkey and tarsier; Fig. 1). The most recently active

[1]Department of Biology, University of Texas, Arlington, Texas 76019, USA. [2]Department of Microbiology, Immunology & Tropical Medicine, George Washington University Medical Center, Washington DC 20037, USA.
*These authors contributed equally to this work.

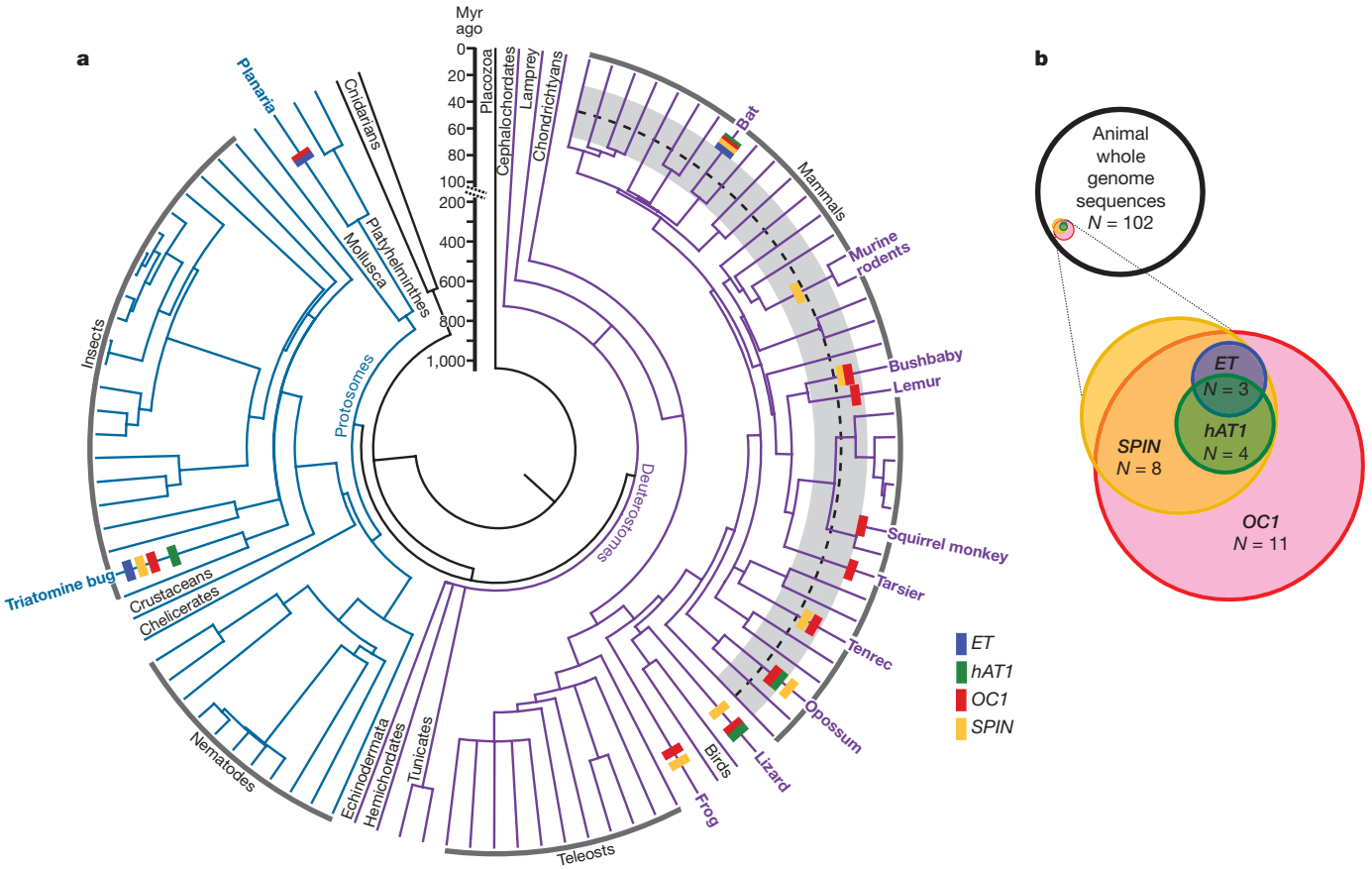

Figure 1 | Taxonomic distribution and age of *SPIN*, *OC1*, *ET* and *hAT1*. **a**, Distribution of the elements among 102 animal genomes for which whole genome sequences are available. Branch lengths of the tree are proportional to time and tick marks illustrate the timing of amplification for each element. The period 60–20 Myr ago is shaded in grey in mammals (dashed line at 40 Myr ago) to facilitate the comparison of the amplification times between species. **b**, Venn diagram illustrating the overlap in taxonomic distribution of the four elements relative to all species for which whole genome sequence data are publicly available.

subfamily of *OC1* was found in bat, although older subfamilies were also apparent in this species, suggesting that transposition activity persisted for a longer period in this lineage or that it was invaded repeatedly (Supplementary Figs 4 and 5). The activity of *hAT1* and *OC1* in different species of mammals appears to have occurred within overlapping evolutionary timeframes (9–26 and 12–34 Myr ago, respectively), which also coincides with the timing of *SPIN* amplification (15–46 Myr ago)[5]. Together, these data suggest that *SPIN*, *hAT1*, *OC1* and *ET* have spread within the past 50 Myr by means of HT among a range of animals spanning four different phyla (Chordata, Arthropoda, Mollusca and Platyhelminthes).

An intriguing aspect of the pattern of HT of these transposons is the widespread geographic distribution of the taxa at the presumed time of the transfers (Fig. 2a). While it is thought that tenrec, lemur, bushbaby and *Xenopus* frogs were restricted to Africa and tarsier to Southeast Asia, there is evidence that the opossum, anole lizard, squirrel monkey and *R. prolixus* were most probably confined to South America (see Methods and references therein). Bats of the genus *Myotis* and muroid rodents have a nearly cosmopolitan distribution, but colonized South America relatively recently (6–10 Myr ago; see Methods), that is, after the HTs had occurred on this continent. Thus, the biogeographical data indicate that HTs of *SPIN* and *OC1* took place on a global scale, occurring on no less than three continents (Asia, Africa and South America). Because South America and Africa broke apart in the Cretaceous more than 65 Myr before the HTs described here, and South America remained an island continent until the formation of the isthmus of Panama in the Pliocene (3–3.5 Myr ago)[13], the taxonomic distribution also implies at least one transoceanic movement of *SPIN* and *OC1* transposons.

Phylogenetic analyses using the consensus sequences of the two most widely distributed transposons (*SPIN* and *OC1*) produced trees with very similar topologies consisting of two strongly supported clusters: one grouping the elements from the South American mammals (opossum and squirrel monkey) with those found in the triatomine bug (endemic to South America) and one grouping all the other species (Fig. 2b, c). This clustering was supported by additional molecular signatures, including a distinctive region in the sequence of *OC1* elements shared by the opossum, squirrel monkey and *R. prolixus* that differentiates them from all the other species (Fig. 2d). Thus, the phylogenetic relationships of the transposon sequences are discordant with those of their host species, but consistent with their continental origin. It is also noteworthy that, whereas in Old World species *OC1* amplification occurred after that of *SPIN*, this order is reversed in New World species where *OC1* appears to have amplified before *SPIN* (Fig. 1). Together, these data not only reinforce the idea that *SPIN* and *OC1* must have travelled between the Old and New World, but also indicate that these elements have spread horizontally among New and Old World taxa within a relatively narrow timeframe (12–46 Myr ago).

Another puzzling aspect of the horizontally transferred transposons is the extensive overlap in their taxonomic distribution. Mapping the occurrence of these transposons onto the phylogeny of 102 animals for which whole genome assemblies are currently available reveals a strikingly non-random distribution, with four species (bat, opossum, anole lizard and *R. prolixus*) sharing all three transposon families (*SPIN*, *OC1*, *hAT1*; Fig. 1, Table 1 and Supplementary Table 4). The probability of observing this distribution by chance alone, if the HT events were independent, is 4.9×10^{-8} (Methods),

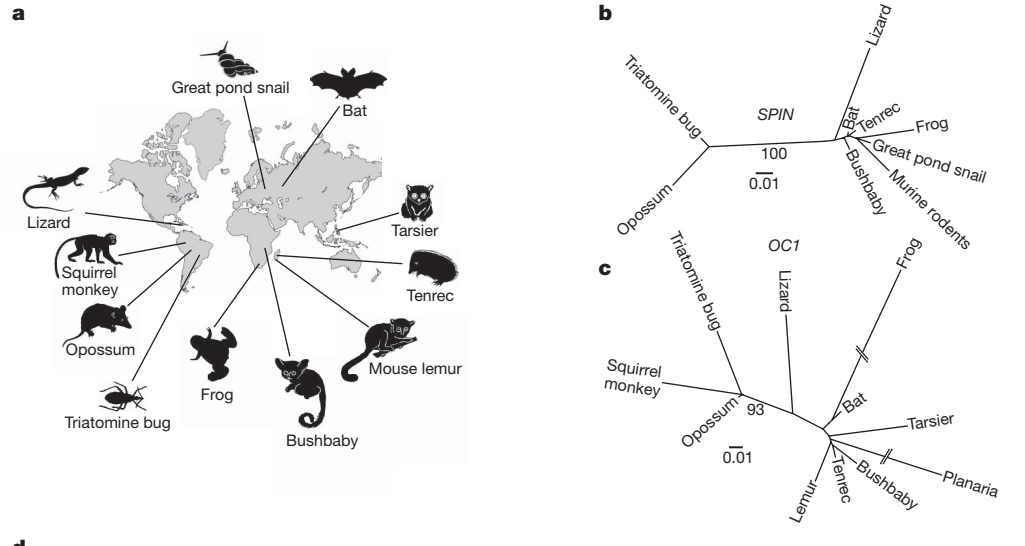

Figure 2 | Biogeographic and phylogenetic evidence supporting horizontal transfers of *SPIN* and *OC1* transposons on multiple continents. **a**, Map showing the likely geographic distribution of the taxa that contain these elements at the time of the transfers (Methods). **b**, **c**, Maximum-likelihood phylogenies of *SPIN* and *OC1* consensuses, respectively; bootstrap values above 70 are shown. **d**, Alignment of representative autonomous (Auto) and non-autonomous (NA) *OC1* consensus sequences, showing portions of the highly conserved 5′ and 3′ termini (black) and the 5′ subterminal region in Old World (light grey) and New World (dark grey) species. All nucleotide positions are with reference to the bushbaby consensus.

suggesting that some ecological factors make these species prone to exchanging genetic material. Among New World taxa, the similarity of *SPIN* and *OC1* consensus sequences of *R. prolixus* with those of opossum and squirrel monkey is striking (95.4–98.1%; Supplementary Table 2). We contend that this reflects HT of these transposons between triatomine bugs and one or more of their mammalian hosts. Triatomine bugs are known to feed on the blood of a variety of mammals in South America, including opossums, squirrel monkeys and bats[14,15]. The exchange of large quantities of blood and saliva between the bugs and their hosts during feeding is known to facilitate the spread of trypanosomes and could also provide a route for the HT of transposons, possibly via these or other intracellular microparasites. Indeed, there is growing evidence for the exchange of genetic material between trypanosomes and their vertebrate hosts[16].

Table 1 | Characteristics of *OC1*, *SPIN*, *hAT1* and *ET* elements

	TE	Copy number	kb of DNA	Avg. distance to consensus (%)
Squirrel monkey	*OC1*	>344	>64	4.1
Tarsier	*OC1*	519	278	4.0
Lemur	*OC1*	12,055	2,278	8.5
Bushbaby	*SPIN*	33,123	3,755	10
	OC1	14,687	2,709	10.6
Bat	*SPIN*	50,319	10,725	3.4
	OC1	7,190	1,440	2.2
	hAT1	19,600	4,664	2.5
	ET	17,461	3,908	4.7
Tenrec	*SPIN*	99,338	16,233	10.5
	OC1	6,837	1,584	8.1
Opossum	*SPIN*	4,807	1,429	5
	OC1	6,025	3,098	10.5
	hAT1	2,666	1,377	8.8
Lizard	*SPIN*	12,138	373	9.3
	OC1	3,902	2,626	5.5
	hAT1	298	161	4.7
Frog	*SPIN*	3,992	742	5.2
	OC1	950	445	8.9
Triatomine bug	*SPIN*	211	75	3.6
	OC1	1,286	495	6.4
	hAT1	552	125	10.9
	ET	131	90	1.8
Planaria	*OC1*	46	267	8.6
	ET	141	70	9.0

Numbers for *SPIN* elements are taken from ref. 5 except for the triatomine bug. Characteristics for each subfamily of *OC1*, *SPIN*, *hAT1* and *ET* are given in Supplementary Table 4.

Among Old World taxa, the *SPIN* phylogeny (Fig. 2b) coupled to the extremely high sequence identity between *L. stagnalis* and the tetrapod taxa (96–98.5%; Supplementary Table 2) are suggestive of HT between snail and tetrapod(s). This transfer could be the result of another parasitic relationship, because *L. stagnalis* is an intermediate host for diverse trematode worms that complete their life cycle in a wide range of vertebrate hosts[17,18]. So far, we have been unable to detect any of the horizontally transferred transposons in the sequenced strains of *Trypanosoma cruzi*, one of the trypanosomes infecting *R. prolixus*, or in *Fasciola hepatica*, a mammalian trematode known to use *L. stagnalis* as an intermediate host. However, the streamlined and fast-evolving genomes of such microparasites might prevent the fixation or preservation of transposons in their genomes. Alternatively, HT might not require chromosomal integration in these species, but could involve extrachromosomal vector(s) such as viruses[19–22].

Our findings suggest that HT of genetic material among animals has occurred on a broader scale than previously appreciated, including four families of DNA transposons and spanning four different animal phyla. Although parasitism has been implicated previously to explain HT on smaller scales[3,23–27], to our knowledge this is the first report of repeated HTs between invertebrates involved in host–parasite interactions and their vertebrate hosts. Although the evolutionary consequences of the transfers described here require further investigation, the sheer amount of DNA generated by the amplification of the transposons (Table 1 and Supplementary Table 4) and the myriad

ways through which mobile elements can alter the structure and function of genomes[28,29] supports the idea that the exchange of genetic material between host and parasite species could strongly affect genome evolution.

METHODS SUMMARY

BLASTN was used to screen all GenBank databases for the presence of *OC1*, *SPIN*, *hAT1* and *ET* transposons. A transposon family was considered to be present in a genome if the reconstructed consensus was at least 85% similar to a known transposon over 80% of its length. A total of 56 consensus sequences were constructed based on alignments of at least ten individual copies using a majority rule (Supplementary Data set 1). Copy number and percentage divergence for each transposon family were determined using these consensus sequences to mask the various genomes with RepeatMasker v. 3.2.7[11]. Estimates of the timing of amplification for each transposon family in each species were derived by dividing the average percentage Jukes-Cantor distance by the neutral mutation rate of the species[5]. Because no reliable neutral mutation rate is available for lizard, frog, triatomine bug and planarian, we used the average mammalian neutral rate[30] as an approximate estimate for the timing of amplification in these species (Fig. 1). Maximum-likelihood phylogenies were carried out with the HKY+G and HKY+I models for *SPIN* and for *OC1* elements, respectively. To verify the presence of the various transposons in all species where they were found computationally we used PCR/cloning/sequencing. To rule out DNA contamination in the two species associated with parasitic life cycles (*R. prolixus* and *L. stagnalis*), we performed PCR using a pair of degenerate primers designed for *rag-1* (a jawed vertebrate-specific gene), with human and opossum DNA as positive controls.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 November 2009; accepted 18 February 2010.

1. Keeling, P. J. & Palmer, J. D. Horizontal transfer in eukaryotic evolution. *Nature Rev. Genet.* **9,** 605–618 (2008).
2. Gladyshev, E. A., Meselson, M. & Arkhipova, I. R. Massive horizontal gene transfer in bdelloid rotifers. *Science* **320,** 1210–1213 (2008).
3. Loreto, E. L., Cararetto, C. M. & Capy, P. Revisiting horizontal transfer of transposable elements in *Drosophila. Heredity* **100,** 545–554 (2008).
4. Diao, X., Freeling, M. & Lisch, D. Horizontal transfer of a plant transposon. *PLoS Biol.* **4,** e5 (2006).
5. Pace, J. K., Gilbert, C., Clark, M. S. & Feschotte, C. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc. Natl Acad. Sci. USA* **105,** 17023–17028 (2008).
6. Novick, P. *et al.* Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. *Gene* **449,** 85–94 (2009).
7. Houck, M. A., Clark, J. B., Peterson, K. R. & Kidwell, M. G. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis. Science* **6,** 1125–1129 (1991).
8. Lent, H. & Wygodzinsky, P. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas disease. *Bull. Am. Mus. Nat. Hist.* **163,** 123–520 (1979).
9. Loy, C. & Haas, W. Prevalence of cercariae from *Lymnaea stagnalis* snails in a pond system in Southern Germany. *Parasitol. Res.* **87,** 878–882 (2001).
10. Ray, D. A., Pagan, H. J. T., Thompson, M. L. & Stevens, R. D. Bats with *hAT*s: Evidence for recent DNA transposon activity in genus *Myotis. Mol. Biol. Evol.* **24,** 632–639 (2007).
11. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-3.0 (1996–2004); ⟨http://www.repeatmasker.org⟩.
12. Robertson, H. M. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 1093–1110 (ASM, 2002).
13. Marshall, E. *et al.* Calibration of the great American interchange. *Science* **204,** 272–279 (1979).
14. Maia Da Silva, F. *et al.* Comparative phylogeography of *Trypanosoma rangeli* and *Rhodnius* (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. *Mol. Ecol.* **16,** 3361–3373 (2007).
15. Maia Da Silva, F. *et al. Trypanosoma rangeli* isolates of bats from Central Brazil: genotyping and phylogenetic analysis enable description of a new lineage using spliced-leader gene sequences. *Acta Trop.* **109,** 199–207 (2009).
16. Hecht, M. M. *et al.* Inheritance of DNA transferred from American trypanosomes to human hosts. *PLoS ONE* **5,** e9181 (2010).
17. Rogan, M. T. *et al.* The occurrence of the trematode *Plagiorchis muris* in the wood mouse *Apodemus sylvaticus* in North Yorkshire, UK. *J. Helminthol.* **81,** 57–62 (2007).
18. Zbikowska, E. *et al.* Infestation of *Lymnaea stagnalis* by Digenea flukes in the Jeziorak lake. *Parasitol. Res.* **99,** 434–439 (2006).
19. Piskurek, O. & Okada, N. Poxviruses as possible vectors of horizontal transfer of retroposons from reptiles to mammals. *Proc. Natl Acad. Sci. USA* **104,** 12046–12051 (2007).
20. Fraser, M. J., Smith, G. E. & Summers, M. D. Acquisition of host cell DNA sequences by baculoviruses: relationship between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedris viruses. *J. Virol.* **47,** 287–300 (1983).
21. Friesen, P. D. & Nissen, M. S. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. *Mol. Cell. Biol.* **10,** 3067–3077 (1990).
22. Jehle, J. A., Nickel, A., Vlak, J. M. & Backhaus, H. J. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. *J. Mol. Evol.* **46,** 215–224 (1998).
23. Heath, B. D., Butcher, R. D. J., Whitfield, W. G. F. & Hubbard, S. F. Horizontal transfer of Wolbachia between phylogenetically distant insect species by a naturally occurring mechanism. *Curr. Biol.* **9,** 313–316 (1999).
24. Huigens, M. E. *et al.* Infectious parthenogenesis. *Nature* **405,** 178–179 (2000).
25. Yoshiyama, M. *et al.* Possible horizontal transfer of a transposable element from host to parasitoid. *Mol. Biol. Evol.* **18,** 1952–1958 (2001).
26. Mower, J. P., Stefanovic, S., Young, G. J. & Palmer, J. D. Gene transfer from parasitic to host plants. *Nature* **432,** 165–166 (2004).
27. Davis, C. C. & Wurdack, K. J. Host to parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* **305,** 676–678 (2004).
28. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41,** 331–368 (2007).
29. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10,** 691–703 (2009).
30. Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA* **99,** 803–808 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Betrán, J. Demuth, T. Fondon, B. Koskella, J. Meik, E. Pritham, Q. Wang and members of the Feschotte laboratory for comments and suggestions during the preparation of the manuscript; M. Batzer, E. Dotson, S. Goodman, A. Prelat, T. Robinson, A. Ropiquet and the Grosell and Sánchez laboratories for the gifts of tissue samples used in this study; and J. Spieth and The Genome Center at Washington University School of Medicine in St Louis for permission to use the *R. prolixus* assembly before publication. C.F. is funded by the National Institutes of Health and S.S. by the National Science Foundation.

Author Contributions C.G., S.S. and C.F. designed research, performed research, and analysed data. J.K.P. contributed data and perl scripts. P.J.B. contributed reagents and materials. C.G., S.S. and C.F. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.F. (cedric@uta.edu).

METHODS

Identification and copy number estimation of *SPIN*, *OC1*, *hAT1* and *ET* elements. The non-coding region of the mammalian *SPIN* superconsensus[5] was used as a query in BLASTN (v. 2.2.14[31]) searches against the GenBank databases from the National Center for Biotechnology Information (NCBI), excluding the genomes of mammals, *Xenopus tropicalis* and *Anolis carolinensis* (where *SPIN* had been previously identified). The following BLASTN parameters were used: gap existence penalty, 5; gap extension penalty, 2; penalty for nucleotide mismatch, −3; reward for nucleotide match, 2. *SPIN* was considered present in a species if the consensus was at least 85% similar at the nucleotide level to the *SPIN* superconsensus[5] over at least 80% of its length.

In order to identify TEs other than *SPIN* that are shared between *R. prolixus* and vertebrate species, we used the Repbase library[32] of vertebrate TEs as a query to perform a batch BLASTN search on the *R. prolixus* genome using the same parameters as above. Three TE families (*OC1*, *hAT1*, *ET*) were identified that are more than 85% similar to mammalian TEs over more than 80% of their length. The taxonomic distribution of these three TEs was then assessed by BLASTN searches against the animal whole genome shotgun databases from NCBI and consensus sequences for each subfamily of *OC1*, *hAT1* and *ET* were reconstructed in each species based on a multiple alignment of at least 10 individual copies (all consensus sequences are provided in Supplementary Data set 1).

To estimate copy number and average percentage divergence of each TE family, we used these respective consensus sequences to mask all genomes in which they were identified with RepeatMasker v. 3.2.7[11]. All fragments larger than 100 base pairs (bp) were used to estimate copy number and calculate average percentage divergence in all species except *A. carolinensis*, where only fragments that were at least 80% of the length of the consensus were considered because of a high level of fragmentation and the presence of many chimaeric elements in this species. A complete consensus sequence for *OC1_NA_1_Xt*, a frog-specific non-autonomous subfamily, could not be confidently reconstructed owing to uncertainty of its internal region. The copy number for this subfamily was estimated based on counts of the 5′ and 3′ terminal regions, for which a reliable consensus sequence could be reconstructed. We observed that the 5′ region of *hAT1_NA_1_Md* (position 1–386) and that of *hAT1_NA_3_Md* (position 1–275) were about twice as diverged from the consensus sequence than the rest of the element copies, probably representing mutational hotspots. We therefore remasked the opossum genome without these regions in order to calculate the average percentage divergence for these two non-autonomous subfamilies separately.

Given that, among the 102 species surveyed, we found *OC1* in 11 species, *ET* in 4 species, and *SPIN* in 8 species, the probability of finding these three horizontally transferred transposons in the same species, if the HTs occur by chance, is $11/102 \times 4/102 \times 8/102 = 3.3 \times 10^{-4}$. The probability that four of the 102 species share these three transposons was then calculated using a binomial distribution B $(4; 102; 3.3 \times 10^{-4})$.

To test the comprehensiveness of whole genome shotgun sequences in the database, we performed TBLASTN on each of the 102 animal genomes with the *ets* domain from *Aedes aegypti* (GenBank protein accession: XP_001654606, region 443–529). Using this sequence as a query, we obtained 10 hits in *A. aegypti* and at least one hit in 93 of the other genomes (with an *e*-value $<10^{-10}$). This indicates the sequencing coverage of at least 92% of genomes for which whole genome shotgun sequencing exists is confirmed to be sufficient to detect a domain from a low copy number gene. This, in turn, indicates that the sequencing coverage should be sufficient to detect the presence or absence of TEs (which are typically present in high copy number) in most, if not all, cases.

Nomenclature. Some of the *OposCharlie1*, *hAT1* and *ExtraTerrestrial* subfamilies reported here correspond to subfamilies that had previously been identified in some of the species included in this study, but named differently (Supplementary Table 4). For example *OposCharlie1*, first described in the opossum *Monodelphis domestica*[11] was named *HAT2_MD* in ref. 33 and *hAT-HT2_MD* in ref. 6. We note that an element that does not correspond to *OposCharlie1* has been named *hAT2_Ml* in the bat *Myotis lucifugus*, where *OposCharlie1* is also found[34]. To avoid confusion, we chose to use the first introduced name for this family: *OposCharlie1*. Also, we note that a non-autonomous subfamily of *ExtraTerrestrial* was identified in the bat *M. lucifugus* and was named *Myotis_nhAT3*[10]. We have now identified the autonomous element from which this non-autonomous element derives, and shown that it was not restricted to *Myotis* but was also present in *R. prolixus* and *S. mediterranea*. For these reasons, we decided to introduce the name *ExtraTerrestrial* (or *ET*) for this family. Lastly, *hAT1* was first described in *M. lucifugus*[10], and its name does not pose any particular problem, so we adopted it herein.

Timing of amplification of each TE family. Estimates of the timing of amplification of each TE family in each species were calculated by dividing the average percentage divergence of each TE family, to which the Jukes and Cantor correction[35] was applied, by the neutral mutation rate of the different species. We used the neutral mutation rates calculated for bushbaby (2.9590×10^{-9}), murine rodents (3.5411×10^{-9}), tenrec (2.9173×10^{-9}), opossum (3.2113×10^{-9}) and bat (2.6920×10^{-9}) in ref. 5. Because no reliable neutral mutation rate is available for lizard, triatomine bug, squirrel monkey, planaria and African clawed frog, we used the average mammalian rate (2.2×10^{-9})[30] to generate timing estimates for these species for illustrative purposes only (Fig. 1). The tree in Fig. 1 includes all 102 animals for which a complete or draft genome assembly is available in the whole genome shotgun database of NCBI (http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi). The tree was written in Newick format and drawn as a circular tree with branch length proportional to time in MEGA 4.0[36] (Supplementary Fig. 6). Most phylogenetic relationships and divergence times are taken from the Timetree of life website (http://www.timetree.org/)[37] except for those within teleosts[38,39], *Drosophila*[40], within nematodes[40,41], the genus *Schistosoma*[42], and for the *Drosophila*/mosquito split[43]. Divergence times between *Rhodnius* and *Acyrtosiphon* and between *Nasonia and Apis* are not known, but we placed both of them at 100 Myr ago for illustrative purposes.

Biogeographical data. The geographical distributions of the taxa where *SPIN*, *OC1*, *hAT1* and *ET* were identified are taken from the literature. Tenrec, bushbaby, mouse lemur and African clawed frog are endemic to Africa, and are believed to have been extant there at the time when the *SPIN* and *OC1* invasions occurred (40–20 Myr ago[44–47]). The Philippine tarsier and all extant Tarsiidae are endemic to several islands of Southeast Asia, and the fossil record indicates that the tarsiid lineage was already restricted to Southeast Asia when it was invaded by *OC1* approximately 20 Myr ago[48,49]. The opossum and all extant and extinct didelphoid marsupials are only known from the New World, and the mouse-sized oppossums, which includes *M. domestica*, have been restricted to South America for at least 40 Myr (ref. 50). Anole lizards are endemic to Central/South America[51]. Bats of the genus *Myotis* originated in Eurasia and now, like muroid rodents, have a nearly cosmopolitan distribution, but they did not disperse to South America until 6–10 Myr ago[52,53]. The squirrel monkey is a platyrrhine or New World monkey, all of which diversified in South America less than 30 Myr ago[54]. There is good evidence supporting South America as the sole centre of origin and diversification of triatomine bugs, including *R. prolixus*[55]. Finally, although the distribution of *L. stagnalis* is at present holartic, the American populations are believed to be a recent introduction from a Eurasian stock[56].

Phylogenetic analyses. Maximum-likelihood (ML) phylogenies of *SPIN* and *OC1* elements were built using PHYML v. 3[57]. Alignments were constructed manually in BioEdit[58] and ambiguous regions were removed (Supplementary Data sets 2 and 3). Nucleotide substitution models were chosen using the AIC criterion in Modelltest[59] (HKY+G for *SPIN* and HKY+I for *OC1*). In order to determine the phylogenetic relatedness of the four *hAT* transposons included in this study, we constructed an amino-acid alignment including their transposase region and that of various other *hAT* families taken from Repbase[32]. We then conducted ML analyses with PHYML v. 3 using the JTT model of amino-acid substitution[60]. The robustness of the nodes was evaluated for all phylogenies by performing a bootstrap analysis involving 1,000 pseudoreplicates of the original matrix.

Testing for purifying selection. To examine the pattern of evolution of *OC1*, *hAT1* and *ET* elements within a particular genome after horizontal transfer, d*N*/d*S* analyses were performed as follows: 50 full-length *OC1* copies were extracted from the opossum genome and all copies of *ET* ($n = 40$) and *hAT1* ($n = 49$) that contained at least 60% of the transposase sequence were retrieved from the bat genome. A multiple alignment of the coding region of these individual copies and their respective consensus was constructed using BioEdit[58] and all non-sense mutations were removed. We then tested whether the pattern of mutations observed between each copy and the consensus (an estimate of the ancestral founder element) was significantly different from what is expected if the sequence is evolving neutrally using the codon-based Z-test in MEGA 4.0[36] with the Nei-Gojobori method and the Jukes-Cantor correction (500 bootstrap replicates). In addition, we used these multiple alignments and an alignment of 50 full-length or nearly full-length *SPIN_NA_12_Rp* (extracted from *R. prolixus*) to construct neighbour-joining phylogenies in MEGA 4.0, with the maximum composite likelihood model and 1,000 bootstrap replicates. See ref. 5 for examples of *SPIN* star-like phylogenies in other taxa.

Similarity plot (Supplementary Fig. 5a). Similarity among copies of autonomous *OC1* elements from the six species in which they were identified (*Otolemur garnettii, Tarsius syrichta, Microcebus murinus, Echinops telfairi, Schmidtea mediterranea* and *Monodelphis domestica*) was calculated using DnaSP v5[61]. Alignments were made using ClustalW[62] and corrected manually. All ambiguous sites were considered fourfold degenerate and were included in the analysis, whereas gapped sites were excluded. Polymorphism (*pi*) was calculated in 10 bp windows using 3 bp stepwise increments over the length of the entire element

(3,291 bp), including the *OC1* transposase which is 1,808 bp long (position 1312–3120 of the bushbaby consensus (Supplementary Data set 1)). Values were converted to percentages and subtracted from 100 to plot similarity. Two species (*S. mediterranea* and *M. domestica*) were unalignable in an 816 bp portion of the element (position 64–880 of the opossum and 64–352 of *S. mediterranea*; Supplementary Data set 1) and were excluded for this region.

PCR/cloning/sequencing. Newly identified *SPIN*, *OC1*, *hAT1* and *ET* elements were validated in each species by PCR amplification, cloning and sequencing. For *R. prolixus*, genomic DNA was extracted from insect legs in order to rule out possible contamination from ingested blood. Primers for each element are listed in Supplementary Table 5. PCR was conducted using the following temperature cycling: initial denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing between 48–54 °C based on element-specific gradients (for 30 s), and elongation at 72 °C for 1 min, ending with a 10 min elongation step at 72 °C. Fragments from the PCR were visualized on a 1–2% agarose gel, cloned and sequenced. Cloning was performed using the Strataclone PCR cloning kit (Stratagene) following manufacturer's protocols and successfully transformed bacterial colonies were screened by PCR (same thermocycling program as above) using M13 primers (see Supplementary Table 5) and gel electrophoresis. Amplicons from cloning products were excised from the gel and soaked in 100 µl double distilled H_2O for 2–4 h. PCR was used to re-amplify the products from this solution (using M13 primers) and sequencing reactions were performed using the reamplified product as template using ABI's BigDye sequencing mix (1.4 µl template PCR product, 0.4 µl BigDye, 2 µl manufacturer supplied buffer, 0.3 µl reverse primer, 6 µl H_2O). The thermocycler program was as follows: 2 min denaturation (96 °C) followed by 30 cycles alternating between 96 °C (30 s) and 60 °C (4 min), ending at 10 °C for 3 min. Sequencing reactions were ethanol precipitated and run on an ABI 3730. Sequences were trimmed using Sequencher 4.8 (Gene Codes) and were aligned and analysed using MEGA 4.0[36]. To further rule out contamination, degenerate primers were designed to amplify *rag-1*, a gene found only among jawed vertebrates, and PCR was performed on DNA extracted from *R. prolixus* and *L. stagnalis* to ensure non-amplification (with human and opossum DNA as a positive control). The thermocycler program for this PCR amplification was the same as that described above, but using 56 °C for annealing.

31. Altschul, S. F. *et al.* Basic Local Alignment Search Tool. *J. Mol. Biol.* **215,** 403–410 (1990).
32. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110,** 462–467 (2005).
33. Gentles, A. J. & Jurka, J. HAT2_MD — an autonomous *hAT* transposon from *Monodelphis domestica*; consensus sequence. *Repbase Report* **5,** 10 (2005).
34. Ray, D. A. *et al.* Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* **18,** 717–728 (2008).
35. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21–132 (Academic, 1969).
36. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24,** 1596–1599 (2007).
37. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22,** 2971–2972 (2006).
38. Yamanoue, Y. *et al.* The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. *Genes Genet. Syst.* **81,** 29–39 (2006).
39. Steinke, D., Salzburger, W. & Meyer, A. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J. Mol. Evol.* **62,** 772–784 (2006).
40. Cutter, A. D. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol. Biol. Evol.* **25,** 778–786 (2008).
41. De Giorgi, C. *et al.* Structural and evolutionary analysis of the ribosomal genes of the parasitic nematode *Meloidogyne artiellia* suggests its ancient origin. *Mol. Biochem. Parasitol.* **124,** 91–94 (2002).
42. Morgan, J. A. T. *et al. Schistosoma mansoni* and *Biomphalaria*: past history and future trends. *Parasitology* **123,** S211–S228 (2001).
43. Wiegmann, B. M. *et al.* Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* **7,** 34 (2009).
44. Yoder, A. D. *et al.* Ancient origin for Malagasy primates. *Proc. Natl Acad. Sci. USA* **93,** 5122–5126 (1996).
45. Evans, B. J. *et al.* A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implication for polyploidy evolution. *Mol. Phylogenet. Evol.* **33,** 197–213 (2004).
46. Wilson, D. E. & Reeder, D. M. *Mammal Species of the World: A Taxonomic and Geographic Reference* (Smithsonian Institution Press, 2005).
47. Poux, C. *et al.* Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. *Syst. Biol.* **54,** 719–730 (2005).
48. Gebo, D. L. *et al.* The oldest known anthropoid postcranial fossils and the early evolution of higher primates. *Nature* **404,** 276–278 (2000).
49. Rossie, J. B., Ni, X. J. & Beard, K. C. Cranial remains of an Eocene tarsier. *Proc. Natl Acad. Sci. USA* **103,** 4381–4385 (2006).
50. Steiner, C., Tilak, M., Douzery, E. J. P. & Catzeflis, F. M. New DNA data from a transthyretin nuclear intron suggest an Oligocene to Miocene diversification of living South America opossums (Marsupialia: Didelphidae). *Mol. Phylogenet. Evol.* **35,** 363–379 (2005).
51. Roughgarden, J. Ed. *Anolis Lizards of the Caribbean. Ecology, Evolution and Plate Tectonics* (Oxford Univ. Press, 1995).
52. Stadelmann, B., Lin, L. K., Kunz, T. H. & Ruedi, M. Molecular phylogeny of New World *Myotis* (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. *Mol. Phylogenet. Evol.* **43,** 32–48 (2007).
53. Steppan, S. J., Adkins, R. M. & Anderson, J. Phylogeny and divergence-date estimates of rapid radiation in muroid rodents based on multiple nuclear genes. *Syst. Biol.* **53,** 533–553 (2004).
54. Poux, C. *et al.* Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Syst. Biol.* **55,** 228–244 (2006).
55. Gaunt, M. W. & Miles, M. A. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol. Biol. Evol.* **19,** 748–761 (2002).
56. Remigio, E. A. Molecular phylogenetic relationships in the aquatic snail genus *Lymnaea*, the intermediate host of the causative agent of fascioliasis: insights from broader taxon sampling. *Parasitol. Res.* **88,** 687–696 (2002).
57. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
58. Hall, T. BioEdit version 5.0.6. ⟨http://www.mbio.ncsu.edu/BioEdit/bioedit.html/⟩ (2004).
59. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14,** 817–818 (1998).
60. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8,** 275–282 (1992).
61. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–1452 (2009).
62. Thompson, J. D., Higgins, D. G. & Gibson, T. J. 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 (1994).