MBE Advance Access published July 18, 2011

Rampant horizontal transfer of *SPIN* transposons in squamate reptiles

Clément Gilbert, Sharon S. Hernandez, Jaime Flores-Benabib, Eric N. Smith, Cédric Feschotte

Department of Biology, University of Texas, Arlington, TX, USA

Abstract

Transposable elements (TEs) are highly abundant in the genome and capable of mobility, two properties that make them particularly prone to transfer horizontally between organisms. While the impact of horizontal transfer of TEs is well recognized in prokaryotes, the frequency of this phenomenon and its contribution to genome evolution in eukaryotes remain poorly appreciated. Here we provide evidence that a DNA transposon called *SPIN* has colonized the genome of 17 species of reptiles representing nearly every major lineage of squamates, including 14 families of lizards, snakes and amphisbaenians. Slot blot analyses indicate that *SPIN* has amplified to high copy numbers in most of these species, ranging from 2000 to 28,000 copies per haploid genome. In contrast, we could not detect the presence of *SPIN* in any of the turtles (7 species from 7 families) and crocodiles (4 species) examined. Genetic distances between *SPIN* sequences from species belonging to different squamate families are consistently very low (average = 0.1), considering the deep evolutionary divergence of the families investigated (most are >100 million years (MY) diverged). Furthermore these distances fall below inter-familial distances calculated for two genes known to have evolved under strong functional constraint in vertebrates (*RAG1*, average = 0.24 and *C-mos*, average = 0.27). These data, combined with phylogenetic analyses, indicate that the widespread distribution of *SPIN* among squamates is the result of at least 13 independent events of horizontal transfers. Molecular dating and paleobiogeographical data suggest that these transfers took place during the last 50 MY on at least three different continents (North America, South America and Africa). Together these results triple the number of known *SPIN* transfer events among tetrapods, provide evidence for a previously hypothesized trans-oceanic movement of

SPIN transposons during the Cenozoic, and further underscore the role of horizontal transfer in the evolution of vertebrate genomes.

Introduction

Horizontal transfer (HT) is the passage of genetic material between reproductively isolated organisms. This mode of transmission is common in prokaryotes, where it often serves as a source of genomic innovation (Ochman et al 2000; Koonin et al 2001). HT is also increasingly recognized as an important evolutionary force shaping eukaryotic genomes. Most HT events characterized so far in eukaryotes correspond to prokaryote-to-eukaryote, or organelle-to-nucleus gene transfers (see Marcet-Houben & Gabaldon 2009 and Danchin et al 2010 for recent examples, and Keeling & Palmer 2008 and Andersson 2009 for reviews). By comparison, relatively few cases of gene transfers have been described between multicellular eukaryotes (but see Gladyshev et al 2008, Moran & Jarvik 2010 and Slot & Rokas 2011), and the majority of reported HTs between metazoans correspond to transfers of transposable elements (TEs; Schaack et al 2010). TEs are segments of DNA that are able to move between different genomic loci, often duplicating themselves in the process (Craig et al 2002). Two properties of these elements suggest that they may be more likely than genes to transfer horizontally between organisms: they are capable of mobility, and they often represent the single most abundant component of eukaryotic genomes–for example, TEs make up ~45% and ~85% of the human and maize genomes respectively (Lander et al 2001; Schnable et al 2009).

Over 200 solid cases of horizontal transfers of transposable elements (HTT) have been described so far in multicellular eukaryotes (Loreto et al 2008; Schaack et al 2010). This number may be viewed as surprisingly high given that only few large-scale searches have been conducted. In one of these searches, we computationally screened all eukaryotic genomes available in GenBank for the presence of

four families of DNA-mediated (or class 2) transposons named *SPACE INVADERS* (or *SPIN*), *OposCharlie1*, *hAT1* and *ExtraTerrestrial* (Pace et al 2008; Gilbert et al 2010). We found that these TEs were patchily distributed among metazoans, and that they exhibited extremely high levels of nucleotide similarity between the different species in which they were found (up to 98%). Given the absence of purifying selection acting on the TE coding sequences following their genomic insertion and the deep evolutionary divergence times separating the various host species (80 to >500 MY), we inferred that such levels of sequence similarity were incompatible with vertical inheritance of the TEs from a common metazoan ancestor. We concluded that the observed taxonomic distribution of these four families of TEs was the result of multiple events of HT. Estimates of the ages of these transposon families suggested that most of the transfers occurred within a relatively narrow evolutionary window, between 50 and 10 MY ago (MYA) (Gilbert et al 2010). Together with other recent studies (Novick et al 2010; Thomas et al 2010; Pagan et al 2010), these results indicate that the frequency and impact of HTT in metazoans may be underappreciated.

One limitation of the aforementioned studies is that they were based on computational analyses of relatively few genome sequences available in the databases. Although the number of eukaryotic genome sequences is increasing rapidly, it still accounts for a minuscule fraction of known extant species (less than 1000 out of >>1 million eukaryotes), and the distribution of sequenced genomes among taxa remains heavily biased towards species closely related to model organisms (Schaack et al 2010). One way to improve our appreciation of the frequency of HTT and of its impact on eukaryotic genomes is to use a targeted and systematic experimental approach in order to extend the search for TEs known to have horizontally transferred to large taxonomic groups for which few or no complete genomes are available. Here we initiate such a strategy by screening a comprehensive taxonomic sample of non-avian reptiles for the presence of *SPIN* transposons using a combination of PCR/sequencing and slot blot

hybridizations. Non-avian reptiles include four major lineages: the squamates (8200 species; Hedges & Vidal 2009), the turtles (313 species; Shaffer 2009), the crocodiles (23 species; Brochu 2003), and the sphenodontians (1 species; Hay et al 2010). The rationale for targeting this taxonomic group was twofold. First we knew that *SPIN* has invaded the green anole lizard *Anolis carolinensis* (Pace et al 2008), the only non-avian reptile that has its genome completely sequenced, indicating that squamates are not refractory to HTT. Second, we have access to a diverse and representative collection of tissue samples through the UT Arlington Amphibian and Reptile Diversity Research Center.

Material and methods

DNA extractions, PCR, cloning, sequencing

The list of tissue samples used in this study is provided in Supplementary Table 1. DNA was extracted using the DNeasy blood and tissue kit (Qiagen). The PCR screening for the presence of *SPIN* was conducted using two different sets of primers. The first one, designed to preferentially amplify small, non-autonomous elements, included one primer anchored in the 5' non coding region of *SPIN* (NAF2 5'-CGA ACG ACC CTT TCA CAG G; position 41 to 59 of the *SPIN* superconsensus provided in Pace et al (2008) and one primer anchored in the 3' non coding region of *SPIN* (NAR2 5'-CAG TTC CTC ATG TTG TGG TGA C; position 2812 to 2833 of the *SPIN* superconsensus). The second set of primers was designed to amplify a ~400-bp fragment of the *SPIN* transposase (*SPIN*tpaseF2 5'-CAT GTT GCC TAC CTT ATC TGC; position 2005 to 2025 of *SPIN* superconsensus and *SPIN*tpaseR2 5'-ACT TGA TAA CCA ACA AGC TGG; position 2379 to 2399 of *SPIN* superconsensus). PCR reactions were 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 at 54 °C (for both primer sets) 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 (M13F 5'-GTA AAA CGA CGG CCA G; M13R 5'-CAG GAA ACA GCT ATG AC; Annealing temperature = 58 °C). Between 5 and 10 amplicons from each cloning products were directly sequenced using ABI's BigDye sequencing mix (1.4 ul template PCR product, 0.4 ul BigDye, 2 ul manufacturer supplied buffer, 0.3 ul reverse primer, 6 ul H2O). 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 sequencer.

The identity of the species included in this study was verified by sequencing a portion of *C-mos*, a nuclear gene that has already been sequenced for many reptile species as part of several phylogenetic studies. PCR reactions for this gene were carried out using the primers and the protocol described in Saint et al (1998). When *C-mos* was not available in public databases (five species, see Supplementary Table 1), we sequenced a portion of the mitochondrial cytochrome oxidase subunit I (*COI*) using the following primers: (*COI*tF 5'-CAC CAG ATA TAG CAT TTC CAC G and *COI*tR 5'-GCT GGG GAT TTT ATG TTG ATT G for turtles and *COI*cF 5'-GCA CCC GAC ATA GCA TTY CC and *COI*cR 5'-TGG GTG GCC GAA GAA TCA G for crocodiles). PCR products were sequenced directly (without cloning), following the protocol described above. All *C-mos* and *COI* sequences produced in this study have been deposited in GenBank (accession numbers: JN090128-JN090158; see also Supplementary Datasets 1 and 2) and SPIN sequences are provided in Supplementary Datasets 3, 4, 5 and 6.

Slot blot

For each species included on the slot blot, 1 ug of genomic DNA was denatured in a solution containing 0.4M NaOH and 10mM EDTA at 90 °C for 10 mins, and deposited onto a Hybond-N+ membrane (Amersham) using a Minifold I Slot-Blot System (Whatman). The concentration of the DNA samples was

measured using a NanoDrop (1000 Spectrophotometer) and by running an Ethidium Bromide stained agarose gel. We noticed a few discrepancies between the DNA concentration measured with the NanoDrop and the relative intensity of the DNA samples on the agarose gel. In these instances, we chose to use the results of the Ethidium Bromide staining to adjust the amount of DNA deposited on the membrane and to ensure that these amounts were equal for all taxa. The membrane was then rinsed in 2X SSC, air dried for 3 mins, and UV cross-linked. In our previous reports based on computational scanning of *SPIN* in whole genome sequences (Pace et al 2008; Gilbert et al 2010), we have observed that the copy numbers of both non-autonomous (NA) and autonomous (auto) *SPIN* can vary greatly between species. For example, there are 4812 *SPIN* copies (4807 NA; < 5 auto) in the opossum (*Monodelphis domestica*) genome sequence and 33,123 *SPIN* copies (25,978 NA; 7,145 auto) in bushbaby (*Otolmeur garnettii*). In addition, we noticed that in some species such as the lizard *Anolis carolinensis*, *SPIN* elements may be highly fragmented. Thus using a single short fragment of the *SPIN* autonomous sequence as a probe for the slot blot would only allow detecting a very minor portion of the total number of *SPIN* copies present in this genome. In order to enhance our ability to detect both autonomous and non-autonomous elements as well as fragmented copies, we used a mixture of three *SPIN* fragments as probe for hybridization. The first fragment is 260 bp long and corresponds to the very 5' end of a *SPIN* copy, amplified in the genome of the tenrec (*Echinops telfairi*) using a primer anchored in the 5' flanking region of this copy (5'dotblot2Fwd 5'-CTT GCC TAA TGT CTT AGA GCA G) as well as a primer anchored at position (221 – 244) of the *SPIN* superconsensus (5'dotblot2Rev 5'-TGT CAC ATA CAC CTG TAT TTG TAC). The second fragment is 304 bp long and corresponds to the very 3' end of the same *SPIN* copy amplified in the tenrec genome using a primer anchored in the 3' flanking region of this copy (3'dotblot54Rev 5'-GTT GGA AGA AAG GTC TAG GTC AG) and a primer anchored at position (2583 – 2603) of the *SPIN* superconsensus (3'dotblot55Fwd 5'-TGC TAC ACC ATG CTT CAA GAC). The third fragment corresponds to the portion of the transposase gene and was amplified in two different

genomes (*Ameiva undulata*, *Sceloporus adleri*) using the primers *SPIN*tpaseF2/R2 (see above for sequence). The PCR products of each of these fragments were cloned as described above. The products obtained from a second PCR performed on one clone for each of the fragments were then mixed and [α-32P]dCTP-labeled using the Random Primed DNA Labeling Kit (Roche). To verify that the transfer of DNA onto the nylon membrane was successful, we used as positive control a 242-bp fragment of the gene encoding the *28S* ribosomal RNA, amplified in *Cordylus tropidosternum* with the following primer set (*28S*F 5'-AGG TGT CCT AAG GCG AGC; *28S*R 5'-GAT AGG AAG AGC CGA CAT CG). This fragment was cloned and labeled as described above. The membrane was first hybridized with the *28S* probe, washed and exposed during 24 hours on X-ray film (Kodak). It was then stripped, exposed for 24 hours to verify that no DNA was left on the membrane after stripping, and hybridized with the *SPIN* probe. After washing the excess of *SPIN* probe, the membrane was exposed for another 24 hours on X-ray film. Hybridizations were performed in PerfectHyb Plus hybridization buffer (Sigma) at 65 °C. For each washing, the membrane was soaked twice 30 mins into a solution of 1X SSC/0.1% SDS.

To approximate the number of *SPIN* copies present in squamate genomes, three dilutions (2.5ng, 1ng and 0.5ng) of unlabeled probe were blotted onto the membrane. The number of *SPIN* fragments contained in each dilution (7.41×10^9, 3.095×10^9 and 1.5475×10^9 respectively) was calculated using the formula "copy number = (amount in ng * number/mole) / (length in bp * ng/g * g/ mole of bp)" described and implemented on the website http://www.uri.edu/research/gsc/resources/cndna.html. The integrated density of the signal produced by each dilution on the dot blot was measured using ImageJ 1.43u (http://rsb.info.nih.gov/ij/). Based on the comparison of these three densities, we calculated that doubling the probe quantity led to a 1.25 - 1.3 fold increase of the signal. We therefore estimated the total copy numbers for each species by comparing the intensity of each signal to that of the signal obtained with the 1ng dilution, and assuming that a 1.27-fold change in density translates into a 2-fold

change in copy number. In order to calculate the copy numbers per haploid genomes, we used C-values taken from the Animal Genome Size Database (Gregory 2011) and assumed that 1 pg = 1 Gb. When multiple C-values were given for one species, we used the average of these values. When a C-value was not available for a species but was available for various other species of the same genus, we used the average of these values. When a C-value was not available for any species belonging to the same genus as one of the species included in this study, we used an arbitrary C-value of 2, corresponding to the average of all other squamate C-values that we were able to find in the database. These genome sizes allowed us to estimate the number of haploid genomes present in 1ug of DNA for each species using the formula given above.

Phylogenetic analyses

All autonomous reptile *SPIN* sequences obtained in this study were aligned together with five *SPIN* sequences randomly extracted from the genome of the species where these elements have been previously described (Pace et al 2008; Novick et al 2009; Gilbert et al 2010). The multiple alignment is provided in Supplementary Dataset 6. Because of the low copy number of autonomous *SPIN* elements in the lizard (*Anolis carolinensis*) and opossum (*Monodelphis domestica*) genome, only one sequence corresponding to the aligned region could be included for these two species. Similarly, all *SPIN* clones sequenced from *Scincus scincus* were identical and we therefore included only one of them in the phylogenetic analysis. Sequences were aligned by hand using BioEdit 7.0.5.3 (Hall 2004) and regions absent in more than half of the sequences were removed. The model of nucleotide evolution best fitting this alignment (TPM3uf+G) was chosen based on the Akaike Information Criterion (AIK) in jModeltest 0.1.1 (Posada 2008). This model was then entered in PhyML 3.0 (Guindon & Gascuel 2003) to perform a bootstrap analysis involving 1,000 replicates in a maximum likelihood framework.

Distances between *SPIN*, *C-mos* and *RAG1* sequences

Pairwise distances between the different reptile species included in this study were calculated for *SPIN*, *C-mos* and *RAG1* sequences. For *SPIN*, the distances were first calculated between majority rule *SPIN* consensus sequences derived for each species based on an alignment of individual copies (except in *Scincus scincus* for which we were able to isolate a single sequence). We also calculated the average distances between individual SPIN copies within each species where multiple SPIN copies were sequenced. For *C-mos*, distances were calculated between the sequences produced during this study (see above) and for RAG-1, we used sequences available in GenBank for the species in which we found *SPIN*, or for closely related species (See Supplementary Table 1 for accession numbers). The model of nucleotide evolution best fitting each alignment was chosen using the AIC in jModeltest (Posada 2008) and was entered in PAUP 4.0 (Swofford 2003) to calculate pairwise distances under maximum likelihood settings. The multiple alignments used to calculate these distances are provided in Supplementary Datasets 1, 2, 3 and 4.

Analyses of selection

For each squamate species where we found *SPIN* elements, we 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 4.036 with the Nei-Gojobori method and the Jukes-Cantor correction (500 bootstrap replicates). The results of the test showed that the number of non-synonymous substitutions accumulated on *SPIN* copies since their insertion in the various squamates was not significantly different from the number of synonymous substitutions. This result indicates that the *SPIN* copies are evolving neutrally after insertion in the genome.

Paleobiogeographic data

In order to infer the most likely distribution of squamate species at the time of the *SPIN* transfers, we gathered fossil and molecular dating evidence from the literature (Figure 1). Currently *Amphisbaena alba* is widely distributed in South America and the ancestors of the genus *Amphisbaena* are thought to have originated on this continent >80 MYA (Macey et al 2004). The family Teiidae (whiptails) likely originated in South America during the Paleocene (>55 MYA) and the presence of *Ameiva undulata* in Central America and southern Mexico is probably the result of recent dispersal (Giugliano et al 2007). According to Fuller et al (1998), varanid lizards arose in Asia around 65 MYA and migrated to Africa, where *Varanus exanthematicus* is found today, during the late Tertiary period. The oldest African fossils attributed to the varanid family are found in the early Miocene of Kenya (~18 MYA; Clos 1995). The current distribution of the anguid lizard *Mesaspis moreletii* covers most highlands of Nuclear Central America including southern Mexico. The species is part of the Gerrhonotinae, a subfamily that arose in North America at least 50 MYA (Macey et al 1999). Beaded lizards (Helodermatidae), today distributed in the southwestern USA, Mexico and Guatemala, originated in North America about 35 MYA (Douglas et al 2010). Furthermore, the fossil record indicates that the ancestors of this family occurred in North America during the early Cretaceous (>100 MYA; Cifelli and Nydma 1995; Nydam 2000). The snakes *Hypsiglena torquata* (Dipsadidae) and *Nerodia erythrogaster* (Natricidae) are found today in the southwestern USA and Mexico and southeastern USA, respectively. The two species belong to the superfamily Dipsadoidea, which likely originated in Asia during the Tertiary. Biogeographical scenarios are however equivocal regarding the precise date at which the ancestors of the Dipsadidae and Natricidae first migrated to the New World (Pinou et al 2004). The Dipsadoidea may have entered the New World either at the same time as the Colubridae (16–10 MYA; Vidal et al 2000) or at the same time as the Boidae (48 MYA; Estes and Hutchinson 1980). The three viperid snakes (*Agkistrodon contortrix*, *Crotalus atrox* and *Sistrurus catenatus*) are widely distributed in North America. It is thought that viperid snakes migrated from Asia to the New World through the Bering Land Bridge and the earliest North

American fossil attributed to the family is from the Early Miocene (22 MY old; Douglas et al 2006). Among iguanids, anole lizards and the subfamily Phrynosomatinae (to which *Sceloporus adleri* belongs) are endemic to Central-South America and North America respectively. While the biogeographic history of the family Iguanidae is complex, both fossil evidence and molecular dating suggest that the ancestors of these two taxa arose in the New World more than 50 MYA (Smith 2009; Noonan and Chippindale 2006). *Agama agama* is currently restricted to Africa but it is unclear whether African agamids originated in Africa or Asia and the age of their ancestor has not been precisely determined within the Cenozoic (Macey et al 2000; Okajima and Kumazawa 2010). *Scincus scincus* is currently found in Africa and it is thought that the family Scincidae (skinks) originated in Africa during the Cenozoic more than 23 MYA (Greer 1970; Whiting et al 2003; Carranza et al 2008). *Lepidophyma flavimaculatum* and the 30 other species of the Xantusidae family (night lizards) are all endemic to North and Central America and Cuba, and molecular dating indicates that they arose on this region around 60 MYA (Vicario et al 2003). *Hemidactylus* geckos most likely originated in the Old World but it is unclear where and when exactly (Carranza et al 2006). Cordylid or spinytail lizards (Cordylidae) are all distributed in sub-saharan Africa and their origin dates back to more than 35 MYA on this continent (Stanley et al 2011). In addition, the sister family of Cordylidae (Gerrhosauridae, plated lizards) is also native from Africa and the oldest fossils attributable to the (Cordylidae + Gerrhosauridae) lineage are from the upper cretaceous (>65 MYA) of Madagascar (Krause et al 2003).

Results

Taxonomic sample and PCR screening

Our search for the presence of *SPIN* transposons in reptiles began with an initial PCR screen of 46 squamate species (snake, lizards and amphisbaenians), six species of crocodiles and 23 species of turtles. The two other groups of reptiles (birds and tuatara) were not included in this study but it is worth mentioning that our Blast searches using the *SPIN* superconsensus as a query on the chicken, zebra finch

and turkey genome (the three birds species for which whole genome sequences are available at present [Hillier et al 2004; Warren et al 2010; Dalloul et al 2010]) did not yield any significant hit, suggesting that *SPIN* is absent from these species. We also extended these searches to the sequence of 10 tuatara BAC clones produced by the NIH Intramural Sequencing Center and representing a total of 1.6 Mb. This search did not retrieve any positive hit either but given that these BAC sequences represent only a small fraction of the tuatara genome (C-value = 5 pg [Olmo 1981]), we cannot exclude that *SPIN* is not present in this genome.

Our initial screen yielded positive PCR products for most squamate species (32 out of 46) and negative amplifications for all crocodiles and turtles (not shown). Based on these results, we narrowed down our taxonomic sampling to retain only the species for which we could obtain sufficient good quality DNA in order to repeat PCRs, cloning and sequencing, and to perform a slot blot analysis. In addition, we selected taxa so as to maximize the coverage of the phylogenetic diversity of non-avian reptiles, and we preferentially chose those species for which the gene *C-mos* was available in GenBank in order to facilitate the validation of species ID (see Material and Methods). In the end, we thoroughly screened 20 species representing 18 of the 58 families of squamates (and 14 of the ~32 non-snake families [Hedges & Vidal 2009]), four species representing the two families of crocodiles (Brochu 2003) and seven species representing half of the 14 families of turtles (Shaffer 2009). The results of this screen are summarized in Figure 1. In summary, we were able to amplify and sequence multiple copies of non-autonomous and/or autonomous *SPIN* in 16 out of 20 squamate species. PCRs at annealing temperatures lower than the one used in squamates yielded DNA bands on agarose gel electrophoresis in some species of turtles and crocodiles, but none of them were confirmed as *SPIN* by DNA sequencing.

Ruling out contamination

SPIN sequences previously characterized in other metazoans are typically more than 95% identical between the different species in which they occur and most do not form monospecific phylogenetic groups (Pace et al 2008; Gilbert et al 2010). This means that contamination of genomic DNA would be very difficult to detect in a PCR screen like the one deployed here. Several lines of evidence, however, suggest that the results presented herein are not due to cross DNA or tissue sample contaminations. First, we verified the species ID of all tissue samples used by sequencing the gene *C-mos* or *COI*. For each of the *C-mos* and *COI* sequence obtained, the best Blast hit in GenBank was the expected species, or a species of the same family in cases where the gene of the expected species was not available in Genbank (note: all sequences generated in this study have now been deposited in Genbank). Furthermore, all the validated sequence reads were clean, without double peaks that would have indicated that some of our DNA samples were heavily contaminated. Third, out of the 13 squamate species that yielded positive *SPIN* PCR/sequence and that were included on the slot blot (see below and Suppplementary Figure 1), 10 produced a strong signal, unlikely to result from minor DNA contaminations (see text below for the 3 species that did not produced a strong signal). Fourth, the structure of each non-autonomous element was species-specific (Figure 2 and Supplementary Dataset 5), i.e. we never sequenced the same sub-family of non-autonomous elements in two different species. Non-autonomous class 2 elements are generally amplified from deletion derivatives of autonomous elements that most likely result from aborted double-stranded gap repair following excision (Engels et al 1990). On occasion, a non-TE 'filler' DNA segment may be inserted during the gap repair process, producing a more complex non-autonomous element with an extraneous internal region (Rubin and Levy 1997, Yan et al 1999). All but two subfamilies of non-autonomous elements sequenced in squamate species correspond to such complex non-autonomous elements (Figure 2). This type of rearranged elements appears to be particularly common in the lizard *Anolis carolinensis*, the only squamate for which whole genome sequences are available so far in GenBank (Novick et al 2010).

Interestingly, the non-TE fragments differed in all species both in terms of length and primary sequence. This observation coupled to the fact that the positions of the deletion breakpoints often vary between species (Figure 2) suggests that the different non-autonomous elements originated independently in each squamate lineage where *SPIN* was identified, and therefore provides strong evidence against contamination. We re-extracted DNA and re-sequenced multiple *SPIN* sequences in the two species that produced no signal on the slot blot (*Lepidophyma flavimaculatum*, *Scincus scincus*). In both cases we obtained non-autonomous elements structurally identical to those we had sequenced previously, suggesting that *SPIN* are indeed present in these two species, though most likely at copy numbers that are too low to be detected by slot blot hybridization. Finally, it is worth mentioning that for one of the squamate family included in this study (Viperidae), the presence of *SPIN* was also verified through another independent investigation based on low-coverage (4.5%) high-throughput genome sequencing of the copperhead (*Agkistrodon contortix*) (Castoe et al 2011). Searching the sequence data generated by these investigators, we identified hundreds of *SPIN* sequences (autonomous and non-autonomous, together occupying ~80 kb of DNA) that are more than 90% identical to the *SPIN* superconsensus reconstructed in Pace et al (2008). Among these, we found two *SPIN* fragments identical to the transposase region isolated herein by PCR. Note these two fragments are also 97% identical to the *SPIN* consensus of another pitviper examined, the massassauga rattlesnake (*Sistrurus catenatus*), and all of these sequences cluster phylogenetically (see below) with the *SPIN* copies sequenced from the two pitvipers included in our screening (*S. catenatus* and *Crotalus atrox*).

Evidence for multiple horizontal transfers of *SPIN* in squamates

For most squamate species in which we detected *SPIN* we were able to sequence between three and five segments of transposase coding regions (Supplementary Table 1). All clones (n=8) sequenced in

Scincus scincus were identical, which together with the absence of signal on the slot blot for this species suggests that there might be only one or a few fragmented remnants of autonomous *SPIN* in this species, as in the genome of *Anolis carolinensis* (Pace et al 2008). Nucleotide genetic distances between individual *SPIN* copies within species were extremely low (Supplementary Table 1; average = 0.034; STDEV = 0.02; range = 0.08–0.01). To assess *SPIN* distances between species, we reconstructed a majority-rule *SPIN* consensus for each species and computed pairwise distances between all consensus sequences (Figure 3; Supplementary Table 2). All these comparisons showed distances that are also extremely low average = 0.1; STDEV = 0.06; range = 0.008–0.27). These distances fall within the range of those computed previously for the same region of the autonomous *SPIN* sequence among metazoan species in which *SPIN* is known to have been introduced horizontally (Pace et al 2008; Novick et al 2010; Gilbert et al 2010). As in these previous studies, we found no evidence of purifying selection acting on the coding sequence of *SPIN* in squamates (see Material and Methods). This suggests that, like in other metazoans, *SPIN* elements are not evolving under functional constraint after insertion in squamate genomes. Almost all (113 out of 120) *SPIN* pairwise distances computed herein involve species that diverged from each other more than 100 MYA (Supplementary Table 2). Given these deep divergence times and the absence of purifying selection acting on *SPIN* sequences, the extremely low pairwise *SPIN* distances seem incompatible with a scenario invoking vertical inheritance of these transposons from the ancestor of squamates. Indeed, for most pairwise comparisons (114 out of 120), the distances calculated for *SPIN* are much lower than those computed for *RAG1* (Supplementary Table 2; average = 0.24; STDEV = 0.07; range = 0.01–0.33) and *C-mos* (Supplementary Table 2; average = 0.27; STDEV = 0.09; range = 0.01–0.49), two essential genes that must have evolved under strong functional constraint in virtually all vertebrate lineages (Figure 3). The average difference between pairwise distances calculated for these genes and for *SPIN* are 0.14 for *RAG1* (STDEV = 0.06; range = 0.01–0.23) and 0.18 for *C-mos* (STDEV = 0.06; range = 0.02–0.31). Together, these data strongly suggest that the presence of

SPIN in most of the major squamate lineages examined in this study is the result of independent HT events that occurred after these lineages diverged from each other.

SPIN phylogeny

The general topology of the *SPIN* phylogeny (Figure 4) is unresolved and star-like, indicating that, following insertion, most transposon copies accumulate discrete substitutions at the neutral rate of the species (consistent with the pattern observed for other DNA transposon families, e.g. see Robertson 2002). In several instances (e.g., *Hemidactylus turcicus*, *Lepidophyma flavimaculatum*, *Heloderma horridum*, *Tenrec ecaudatus*, *Otolemur garnettii*) and as previously observed (Pace et al 2008; Gilbert et al 2010), *SPIN* copies do not form monospecific clusters, suggesting that the genome of multiple species were invaded by very similar or identical founder elements. In some species (e.g., *Agama agama*, *Ameiva undulata*, *Nerodia erythrogaster*, *Sceloporus alderi*) however *SPIN* copies cluster into strongly supported monospecific groups separated from the other *SPIN* copies by relatively long branches (Figure 4). This pattern indicates that the founder element invading these species is phylogenetically distinct from the elements identified in other metazoan genomes. There are two instances in the phylogeny where *SPIN* copies from different host species cluster together in a strongly supported clade (color branches in Figure 4). The first cluster includes copies from an amphisbaenian (*Amphisbaena alba*), a marsupial (*Monodelphis domestica*) and a hemipteran insect (*Rhodnius prolixus*). This grouping is not only completely incongruent with the species phylogeny, further supporting horizontal transfer, it may also be indicative of direct transfers between these species given their geographical and ecological overlap (see Discussion section). The second cluster groups *SPIN* copies from the three viperid snakes examined in this study (bootstrap = 65). Given the taxonomic proximity of these snakes, this grouping most likely reflects that *SPIN* was horizontally introduced in a common ancestor of these species (see below).

***SPIN versus RAG1* and *C-mos* distances: exceptions to the trend**

For all pairwise comparisons where *SPIN* distances are lower than the distances for *RAG1* and *C-mos* (126 out of 136 pairwise comparisons), the introduction of *SPIN* via HT must have occurred at a time when *RAG1* and *C-mos* were already very divergent from each other (considering purifying selection), i.e., well after the speciation events separating each pair of species compared. For ten pairs of species however, the distances calculated between *SPIN* consensus sequences are greater than the distances calculated for the two genes (see A and B dotted circles on Figure 3 and Supplementary Table 2). All ten comparisons involve snake species that have diverged from each other less than 55 MYA and are at least 50 MY less divergent than most other species pairs. Two hypotheses may explain the inverse relationship between *SPIN* and *RAG1*/*C-mos* distances for these ten pairs of species: (1) *SPIN* was horizontally transferred in the ancestor of the five species involved in these comparisons, was subsequently transmitted vertically and accumulated mutations faster (at the neutral rate of each species) than *RAG1* and *C-mos* that are evolving under functional constraints, (2) *SPIN* was introduced horizontally in each species but the time at which these HT events occurred was very close to the time at which the species diverged, such that at the time of the transfer, *RAG1* and *C-mos* were still very similar between each two lineages of snakes. Some of these comparisons involve three viperid snakes (*Agkistrodon contortrix*, *Crotalus atrox* and *Sistrurus catenatus*; dotted circle B on Figure 3) that diverged less than 22 MYA (Douglas et al 2006). The *SPIN* copies from these three snakes cluster together in the phylogeny (Figure 4), consistent with the idea that a distinct variant of *SPIN* was horizontally transferred to a common ancestor of these three species. The phylogeny is however inconclusive regarding the relationship of the other snake *SPIN* copies (sequenced from *Nerodia erythrogaster* and *Hypsiglena torquata*) and it is therefore unclear whether these two species each acquired the element *via* HT or whether there was a single transfer event in their common ancestor.

Discussion

In this study we have shown that *SPIN* transposons have infiltrated the germlines of many species of lizards, snakes, and at least one amphisbaenian spreading over most of the diversity of extant squamates (Figure 1). This distribution implies that *SPIN* was transferred independently at least 13 times during the evolution of squamates, which triples the number of HT events of *SPIN* previously reported among diverse metazoans (Pace et al 2008, Novick et al 2010, Gilbert et al 2010). This estimate is conservative as it assumes a single HT event in the common ancestor of all snake species tested positive for *SPIN.* Yet the patchy distribution of *SPIN* among snakes (*SPIN* is absent in the two species of Boidae included in this study) is suggestive of multiple transfers. In any case, these data considerably extend our appreciation of the frequency of HTT in non-avian reptiles, which was previously based on only one sequenced genome, that of the green anole lizard (Gilbert et al 2010; Novick et al 2010; Thomas et al 2010).

The activity of *SPIN* in squamates has resulted in the spread and accumulation of large numbers of elements in a lineage-specific fashion (between 2000 and 28,000 per haploid genome as revealed by slot blot analysis; Figure 1; Supplementary Figure 1; Supplementary Table 1). Further studies will be necessary to assess the evolutionary impact of these invasions on the structure and function of squamate genomes, but it is tantalizing to speculate that they have provided a rich source of raw material for genetic innovation, as observed for other DNA transposons (Feschotte and Pritham 2007). Some recent studies indeed suggest that TEs have been instrumental in the size expansion of the squamate *Hox* gene clusters and that they may have had a major impact on the evolution of body plan in this group (Di-Poï et al 2009, Di-Poï et al 2010). Whether *SPIN* invasions and other HTT events could have been involved in the spectacular diversification of squamates during the Tertiary leading to the

formation of more than 8000 extant species (Hedges & Vidal 2009) is a fascinating question that warrants further investigation.

An interesting aspect of *SPIN* HTs previously noticed is that the transfers apparently took place on at least three different continents (Africa, Eurasia and South America) within a relatively narrow evolutionary window (15–46 MYA) (Pace et al 2008; Gilbert et al 2009; Gilbert et al 2010). One way to estimate the time at which HTT took place is to calculate the pairwise divergence between all individual TE copies and an ancestral founder copy, which can be approximated by the consensus sequence of the TE family. Because TEs typically evolve neutrally after insertion in the genome (e.g., Lampe et al 2003 and see Results section), the time of the transfers can be approximated by dividing these pairwise divergences by the nuclear neutral substitution rate of the various species lineages in which HTT has occurred (Pace et al 2008; Khan et al 2006). The average pairwise corrected divergences between each *SPIN* copy sequenced in this study and the *SPIN* consensus reconstructed for each squamate species ranges between 0.002 and 0.05 (average = 0.019; STDEV = 0.014). Because there is no reliable estimate of neutral substitution rate for non-avian reptiles, we used the two most extreme neutral rates known so far for amniotes (2×10^{-9} and 4.5×10^{-9} substitutions per site per year [subst./site/year]) as a range to estimate the timing of *SPIN* HTs in squamates. These rates were taken from a study of multiple introns in birds (2×10^{-9}-3.9×10^{-9} subst./site/year; Axelsson et al 2004) and on two studies of ancestral orthologous repeats shared between several placental mammals (2.2×10^{-9}-4.5×10^{-9} subst./site/year; Pace et al 2008, Waterston et al 2002). Applying these rates to the squamate *SPIN* pairwise distances yields invasion times ranging from 667 thousand years (using the smallest average divergence [0.002] and the fastest rate [4.5×10^{-9} subst./site/year]) and 25 MY (using the largest average divergence [0.05] and the slowest rate [2×10^{-9} subst./site/year]). These rough estimates indicate that the multiple horizontal introductions of *SPIN* in squamates all occurred within the same evolutionary timeframe (and

some perhaps more recently) as those previously reported in other tetrapods (Pace et al 2008; Gilbert et al 2010).

These estimates should be used cautiously, however, not only because of uncertainties and variations in neutral molecular clocks, but also because we analyzed only few *SPIN* copies for each species and a relatively small segment (~400-bp) of coding sequence. For the sake of the discussion, we assume that all *SPIN* transfers in squamates took place within the past 25 MY, and it is important to note that even if the *SPIN* invasion times were twice older than our oldest estimate (i.e., as old as 50 MY), the following biogeographical inferences would all remain valid. Robust fossil evidence from this period and molecular dating suggests that the ancestors of *Amphisbaena* (Amphisbaenidae), those of *Mesaspis* (Anguidae), *Heloderma* (Helodermatidae) and *Lepidophyma* (Xantusiidae), and those of *Cordylus* (Cordylidae) were distributed on the same three continents occupied today by their extant descendants, namely South America, North America and Africa respectively (Figure 1 and see Material and Methods). Together these observations confirm the hypothesis of at least one transoceanic transfer of *SPIN* elements (See Gilbert et al 2009 for discussion), and they extend the geographic range of *SPIN* HTs to North America. Based on the paleobiogeographical data available for the Varanidae, Agamidae, snakes and *Hemidactylus* over the same period (see Material and Methods), it is possible that *SPIN* HT also took place in Asia (Figure 1). Further sampling of additional *SPIN* sequences and of other species will be required to infer more precise estimates of the time and geographical span of the transfers.

This study, together with earlier reports of HTT on a wide geographical and taxonomic scale (e.g. Pace et al 2008; Gilbert et al 2010), continues to raise the question of the mechanisms underpinning such transfers. A number of possible vectors and scenarios have been proposed in the literature (e.g., Houck et al 1997, Piskurek and Okada 2007; reviewed in Loreto et al 2008 and Schaack et al 2010) but none

have been fully verified in nature. In a recent study we proposed that host-parasite interactions might have facilitated HTT (Gilbert et al 2010). This hypothesis was based on the observation that four families of DNA transposons (including *SPIN*; see introduction) known to have transferred horizontally in multiple species of tetrapods are also found in the kissing bug (*Rhodnius prolixus*; Triatominae), a South-American hemipteran insect feeding on the blood of a wide range of vertebrate hosts. Strikingly two of the transposons (*SPIN* and *OposCharlie1*) identified in *R. prolixus* are almost identical (>98% identity) and cluster phylogenetically with those of the opossum, one of the bug's preferred hosts in nature. This data is consistent with a transfer between host and parasite, an act that could have been facilitated by frequent physical contacts between the bug's saliva and the blood of its hosts when the bug is feeding. The phylogenetic analysis presented herein (Figure 4), although based only on a portion of the *SPIN* autonomous sequence, recovers a strongly supported grouping of *R. prolixus* and opossum *SPIN* sequences. Interestingly, this clade also includes all *SPIN* copies isolated from *Amphisbaena alba*. Amphisbaenians are enigmatic worm-like fossorial squamates found in tropical and semi tropical regions of the world (Macey et al 2004). *Amphisbaena alba* is found throughout most of South America, east of the Andes (Colli and Zamboni 1999), a geographic range that largely overlaps with that of *R. prolixus* and triatomine bugs in general (Lent and Wygodzinsky 1979). There is no published observation of *R. prolixus* feeding specifically on *A. alba*, but numerous triatomine species have been reported to feed on various squamate species (Lent and Wygodzinsky 1979). In addition, triatomines are commonly found under fallen logs, among exposed roots and under loose bark, as well as in the burrows of some mammalian species (Lent and Wygodzinsky 1979). Given the geographical and potential ecological overlap of triatomine bugs and *A. alba* in South America, the hypothesis of a transfer of *SPIN* between these two species seems plausible. These observations further indicate that triatomine bugs might have played a central role in the spread of *SPIN* across diverse South American tetrapods.

Another intriguing trend emerging from this work lies in our inability to detect *SPIN* elements in any of the crocodiles and turtles examined, while we found evidence for their repeated invasions in nearly all major squamate lineages and in several mammals as reported previously (Pace et al 2008). Though our taxonomic sample includes only 7 out of 313 and 4 out of 23 described species of turtles and crocodiles respectively, it does capture a substantial portion of their evolutionary history (Figure 1). Therefore, if HT of *SPIN* occurred at all in these taxa, its frequency is likely to be much lower than in squamates and mammals. The reasons for this difference remain unclear, but it might reflect a lack of (or reduced) interaction between the putative vectors of *SPIN* HT and crocodiles and turtles and/or a lower tolerance or better genomic defense of crocodile and turtle genomes against the amplification of some TEs (against TEs). We note however that retroelements and other DNA transposons have been identified in crocodiles and turtles (Ray et al 2005; Shedlock et al 2007; Kordis 2009). Another simple, non-mutualistic explanation for the absence of *SPIN* HTs in crocodiles and turtles is that these taxa have diversified in a much fewer number of lineages than squamates during the tertiary, leading to a much lower number of extant species (8200 squamates *versus* 313 turtles and 23 crocodiles; Hedges & Vidal 2009; Shaffer 2009; Brochu 2003). Based on these numbers only, the odds of successful HT may have simply been lower in turtles and crocodiles than in squamates. A more systematic screen of crocodiles, turtles and other tetrapods as well as some of their parasites should further our understanding of the taxonomic and geographic spread of these transposons and of the mechanisms and factors promoting HTT and lead to a greater appreciation of the impact of HTT on the evolution of eukaryotic genomes.

Acknowledgements

We are grateful to Andrea Acevedo, Christian Cox, Oscar Flores-Villela, Thomas Eimermacher, Carl Franklin, Alan Kardon (San Antonio Zoo), Jesse Meik, Terry Robinson, Anne Ropiquet, Jeff Streicher, Walter Schargel, and Luis Sigler (Dallas World Aquarium) for the generous gift of tissue samples. In

particular, Carl Franklin was pivotal in providing testudine and crocodilian samples and data. We thank Sarah Schaack and members of the Genome Biology Group at the University of Texas at Arlington for useful suggestions during the preparation of the manuscript. This work was supported by grant R01GM77582 to CF from the National Institutes of Health.

References

Andersson JO. 2009 Gene transfer and diversification of microbial eukaryotes. *Annu Rev Microbiol.* 63:177–193.

Axelsson E, Smith NG, Sundstrom H, Berlin S, Ellegren H. 2004. Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and turkey. *Mol Biol Evol*. 21:1538–1547.

Brochu CA. 2003. Phylogenetic approaches toward crocodylian history. *Annu Rev Earth Planet Sci*. 31:357–97.

Carranza S, Arnold EN. 2006. Systematics, biogeography, and evolution of *Hemidactylus* geckos (Reptilia: Gekkonidae) elucidated using mitochondrial DNA sequences. *Mol Phylogenet Evol*. 38:531–45.

Carranza S, Arnold EN, Geniez P, Roca J, Mateo JA. 2008. Radiation, multiple dispersal and parallelism in the skinks, *Chalcides* and *Sphenops* (Squamata: Scincidae), with comments on *Scincus* and *Scincopus* and the age of the Sahara Desert. *Mol Phylogenet Evol*. 46:1071–94.

Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning AP, Fox SE, Poole AW, Vemulapalli V, Daza JM, Mockler T, Smith EN, Feschotte C, Pollock DD. 2011. Discovery of highly divergent repeat landscapes in snake genomes using high throughput sequencing. *Genome Biol.* (in press)

Cifelli RL, Nydam RL. 1995. Primitive, helodermatid-like platynotan from the early cretaceous of Utah *Herpetologica*. 51:286–291.

Clos LM. 1995. A new species of *Varanus* (Reptilia, Sauria) from the Miocene of Kenya. *J Vertebr Paleontol.* 15:254–267.

Colli GR, Zamboni DS. 1999. Ecology of the worm-lizard *Amphisbaena alba* in the Cerrado of central Brazil. *Copeia*. 3:733–742.

Craig NL, Craigie R, Gellert M, Lambowitz AM. 2002. Mobile DNA II. Washington, DC: American Society for Microbiology Press.

Dalloul RA, Long JA, Zimin AV, et al. (71 co-authors) 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biol*. 8:e1000475.

Danchin EGJ, Rosso MN, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abada P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci USA*. 107:17651–17656.

Di-Poï N, Montoya-Burgos JI, Duboule D. 2009. Atypical relaxation of structural constraints in Hox gene clusters of the green anole lizard. *Genome Res*. 19: 602-10.

Di-Poï N, Montoya-Burgos JI, Miller H, Pourquié O, Milinkovitch MC, Duboule D. 2010. Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature*. 464: 99-103.

Douglas ME, Douglas MR, Schuett GW, Porras LW. 2006. Evolution of rattlesnakes (Viperidae; *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. *Mol Ecol*. 15:3353–3374.

Douglas ME, Douglas MR, Schuett GW, Beck DD, Sullivan BK. 2010. Conservation phylogenetics of helodermatid lizards using multiple molecular markers and a supertree approach. *Mol Phylogenet Evol*. 55:153–167.

Engels WR, Johnsonschlitz DM, Eggleston WB, Sved J. 1990. High-frequency P-element loss in *Drosophila* is homolog dependent. *Cell*. 62:515–525.

Estes R, Hutchinson JH. 1980. Eocene lower vertebrates from Ellesmere Islands, Canadian Arctic Archipelago. *Paleogeogr Paleoclimatol Paleoecol* 30:325–347.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*. 41:331–368.

Fuller S, Baverstock P, King D. 1998. Biogeographic origins of goannas (Varanidae): A molecular perspective. *Mol Phylogenet Evol*. 9:294–307.

Gilbert C, Pace JK, Feschotte C. 2009. Horizontal *SPIN*ning of transposons. *Commun Integr Biol*. 2:117–119.

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*. 464:1347–50.

Giugliano LG, Garcia Collevatti R, Rinaldi Colli G. 2007. Molecular dating and phylogenetic relationships among Teiidae (Squamata) inferred by molecular and morphological data. *Mol Phylogenet Evol*. 45:168–179.

Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid rotifers. *Science*. 320:1210–1213.

Greer AE. 1970. A subfamilial classification of scincid lizards. *Bull Mus Comp Zool*. 139:151–183.

Gregory TR. 2011. Animal Genome Size Database. http://www.genomesize.com.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Hall T. 2004. BioEdit version 5.0.6. Available: http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Hay JM, Sarre SD, Lambert DM, Allendorf FW, Daugherty CH. 2010. Genetic diversity and taxonomy: a reassessment of species designation in tuatara (Sphenodon: Reptilia). *Conserv Genet*. 11:1063–1081.

Hedges SB, Vidal N. 2009. Lizards, snakes and amphisbaenians (Squamata). In Blair Hedges S, Kumar S, editors. The timetree of life. Oxford University Press 2009. p. 383–389.

Hillier LW, Miller W, Birney E, et al. (175 co-authors). 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 432:695–716.

Houck MA, Clark JB, Peterson KR, Kidwell MG. 1991. Possible horizontal transfer of Drosophila genes by the mite *Proctolaelaps regalis*. *Science*. 253: 1125-8.

Jehle JA, Nickel A, Vlak JM, Backhaus H. 1998. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. *J Mol Evol*. 46:215–224.

Keeling PJ, Palmer JD. 2008. Horizontal transfer in eukaryotic evolution. *Nature Rev Genet*. 9:605–618.

Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res*. 16:78–87.

Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu Rev Microbiol*. 55:709–42.

Kordis D. 2009. Transposable elements in reptilian and avian (sauropsida) genomes. *Cytogenet Genome Res*. 127: 94-111.

Krause DW, Evans SE, Gao KQ. 2003. First definitive record of Mesozoic lizards from Madagascar. *J Vertebr Paleontol.* 23:842–856.

Lampe DJ, Witherspoon DJ, Soto-Adames FN, Robertson HM. 2003. Recent horizontal transfer of Mellifera subfamily Mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. *Mol Biol Evol*. 20:554–562.

Lander ES, et al. (255 co-authors). 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Lent H, Wygodzinsky P.1979. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas disease. *Bull Am Mus Nat Hist*. 163:123–520.

Loreto EL, Carareto CM, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity.* 100:545–554.

Macey JR, Schulte II JA, Larson A, Tuniyev BS, Orlov N, Papenfuss TJ. 1999. Molecular phylogenetics, tRNA evolution, and historical biogeography in Anguid lizards and related taxonomic families. *Mol Phylogenet Evol*. 12:250–272.

Macey JR, Schulte JA 2nd, Larson A, Ananjeva NB, Wang Y, Pethiyagoda R, Rastegar-Pouyani N, Papenfuss TJ. 2000. Evaluating trans-tethys migration: an example using acrodont lizard phylogenetics. *Syst Biol.* 49:233–56.

Maceya JR, Papenfuss TJ, Kuehl JV, Fourcade HM, Boore JL. 2004. Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. *Mol Phylogenet Evol*. 33:22–31.

Marcet-Houben M, Gabaldón T. 2009. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet*. 26:5–8.

Moran NA, Jarvik T. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*. 328:624–627.

Noonan BP, Chippindale PT. 2006. Vicariant origin of malagasy reptiles supports late cretaceous antarctic land bridge. *Am Nat*. 168:730–741.

Novick P, Smith J, Ray D, Boissinot S. 2010. Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. *Gene*. 449:85–94.

Nydam RL. 2000. A new taxon of helodermatid-like lizard from the Albian-Cenomanian of Utah. *J Vertebr Paleontol*. 20:285–294.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 405:299–304.

Okajima Y, Kumazawa Y. 2010. Mitochondrial genomes of acrodont lizards: timing of gene rearrangements and phylogenetic and biogeographic implications. *BMC Evol Biol*. 10:141.

Olmo E. 1981. Evolution of genome size and DNA-base composition in reptiles. *Genetica*. 57:39–50.

Pace JK 2nd, Gilbert C, Clark MS, Feschotte C.2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci USA*. 105:17023–17028.

Pagan HJ, Smith JD, Hubley RM, Ray DA. 2010. PiggyBac-ing on a primate genome: novel elements, recent activity and horizontal transfer. *Genome Biol Evol*. 2:293–303.

Pinou T, Vicario S, Marschner M, Caccone A. 2004. Relict snakes of North America and their relationships within Caenophidia, using likelihood-based Bayesian methods on mitochondrial sequences. *Mol Phylogenet Evol*. 32:563–74.

Piskurek O, Okada N. 2007 Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc Natl Acad Sci U S A*. 104: 12046-12051.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 25:1253–1256.

Ray DA, **Hedges DJ**, **Herke SW**, **Fowlkes JD**, **Barnes EW**, **LaVie DK**, **Goodwin LM**, **Densmore LD**, **Batzer MA**. 2005. Chompy: an infestation of MITE-like repetitive elements in the crocodilian genome. ***Gene.*** 362:1–10.

Robertson, H. M. 2002. In Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. ASM. p. 1093–1110.

Roos J, Aggarwal RK, Janke A. 2007. Extended mitogenomic phylogenetic analyses yield new insight into crocodylian evolution and their survival of the Cretaceous–Tertiary boundary. *Mol Phylogenet Evol*. 45:663–673.

Rubin E, Levy AA. 1997. Abortive gap repair: underlying mechanism for Ds element formation. *Mol Cell Biol*. 17:6294–6302.

Saint KM, Austin CC, Donnellan SC, Hutchinson MN. 1998. *C-mos,* a nuclear marker useful for squamate phylogenetic analysis. *Mol Phylogenet Evol*. 10:259–263.

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

Schnable PS, Ware D, Fulton RS, et al. (157 co-authors). 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 326:1112–1115.

Shaffer HB. 2009. Turtles (Testudines). In Blair Hedges S, Kumar S, editors. The timetree of life. Oxford University Press. p. 398–401.

Shedlock AM, Botka CW, Zhao S, Shetty J, Zhang T, Liu JS, Deschavanne PJ, Edwards SV. 2007. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc Natl Acad Sci USA*. 104:2767–2772.

Slot JC, Rokas A. 2011. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Curr Biol.* 21:134–139.

Smith KT. 2009. Eocene lizards of the clade *Geiseltaliellus* from Messel and Geiseltal, Germany, and the early radiation of Iguanidae (Reptilia: Squamata). *Bulletin of the Peabody Museum of Natural History*. 50:219–306.

Stanley EL, Bauer AM, Jackman TR, Branch WR, Mouton , PLeF. 2011. Between a rock and a hard polytomy: Rapid radiation in the rupicolous girdled lizards (Squamata: Cordylidae). *Mol Phylogenet Evol*. 58:53–70.

Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4. Sunderland (Massachusetts): Sinauer.

Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol Evol*. 2:656–664.

Vicario S, Caccone A, Gauthier J. 2003. Xantusiid ‘‘night’’ lizards: a puzzling phylogenetic problem revisited using likelihood-based Bayesian methods on mtDNA sequences. *Mol Phylogenet Evol*. 26:243–261.

Vidal N, Kindl SG, Wong A, Hedges SB. 2000. Phylogenetic relationships of Xenodontine snakes inferred from 12S and 16S ribosomal RNA sequences. *Mol Phylogenet Evol*. 14:389–402.

Warren WC, Clayton DF, Ellegren H, et al. (82 co-authors). 2010. The genome of a songbird. *Nature*. 464:757–762.

Waterston RH, Lindblad-Toh K, Birney E, et al. (222 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 420:520–562.

Whiting AS, Bauer AM, Sites FW Jr. 2003. Phylogenetic relationships and limb loss in sub-Saharan African scincine lizards (Squamata: Scincidae). *Mol Phylogenet Evol*. 29:582–598.

Yan X, Martinez-Fèrez IM, Kavchok S, Dooner HK. 1999. Origination of Ds elements from Ac elements in maize: evidence for rare repair synthesis at the site of Ac excision. *Genetics*. 152:1733–1740.

Figure legends

Figure1. Phylogenetic tree of non-avian reptile taxa screened in this study and overview of the results. The phylogenetic relationships and divergence times are taken from Hedges and Vidal (2009), Shaffer (2009) and Roos et al (2007). The tree summarizes the results of the slot blot and PCR screening. *SPIN* copy numbers are derived from Supplementary Figure 1 (see also Material and Methods and Supplementary Table 1). Also listed is the likely geographic distribution of the taxa during the last ~50 MY (the time during which *SPIN* were transferred horizontally to the various squamate lineages). The references used to infer these ancestral distributions are cited in the Material and Methods section. SA = South America; NA = North America; As. = Asia; Afr. = Africa; NW = New World; OW = Old World; slash bar = or/and (for example, NA/SA indicates that the taxa was present either in North America, either in South America, or on both continent at the time of the transfer). Museum voucher numbers of the tissues, when available, are listed in Supplementary Table 1. The "-?" symbol next to tuatara indicates that we have not found any *SPIN* sequence in the BAC inserts available in NCBI but that given that these sequences only represent a small portion of the tuatara genome (see Results section), we cannot conclude on whether *SPIN* is present or absent in this genome.

Figure 2. Schematic representation of the structure of non-autonomous *SPIN* elements sequenced in various species of squamates. For each non-autonomous element, the 5' and 3' regions that are homologous to the autonomous element are represented in black (the length of each sequence is indicated on top of them in bp). The internal, non-*SPIN* region is represented in grey. All internal non-*SPIN* regions differ in terms of primary sequence (no homology between them) and length (indicated on top of each sequence in bp).

Figure 3. Graph illustrating the pairwise corrected distances of *SPIN* (black lozenges), *RAG1* (grey triangles) and *C-mos* (open circles) between the squamate lineages included in this study. The distances are derived from all possible pairwise comparisons (n = 136; labeled on the X axis) between the 17 species of squamates in which *SPIN* was found. Distances as well as divergence times for each pairwise comparison are listed in Supplementary Table 2. Before plotting the distance values, they were sorted by ascending order of *SPIN* distances. Distances falling into A and B dotted circles correspond to pairwise comparisons for which *RAG1* and *C-mos* distances are lower than *SPIN* distances (see Result section).

Figure 4. Phylogenetic analysis of *SPIN* sequences. The tree is based on an alignment of a ~400-bp region of the *SPIN* transposase. It includes all autonomous squamate *SPIN* copies sequenced in this study and five copies randomly extracted from the genomes of the species in which *SPIN* has been previously characterized. The genome of *Anolis carolinensis* and *Monodelphis domestica* contain only one autonomous *SPIN* copy that includes the transposase region used in this analysis. The phylogenetic analysis was carried out in a maximum likelihood framework using PhyML 3.0. Bootstrap values (BP) were obtained after analysis of 1000 replicates of the original matrix and are shown as open circles (70 < BP < 80), black circles (81 < BP < 90), or asterisks (90 < BP < 100).

Supplementary Figure 1. Slot blot showing the hybridization of genomic DNA extracted from various species of squamates with a radiolabelled *SPIN* probe. The tenrec (*Tenrec ecaudatus*), a species in which *SPIN* is known to have reached high copy number (>90,000) was used as a positive control. Human, where *SPIN* is absent, was used as a negative control. The three probe dilutions used to estimate copy numbers (see Material and Methods) are deposited on the first lane of the blot (*SPIN* on the left, *28S* rRNA on the right). The same membrane was hybridized with a probe corresponding to a fragment of the *28S* rRNA gene in order to verify the presence of DNA on the membrane (Supplementary Figure 2).

Supplementary Figure 2. Slot blot showing the hybridization of genomic DNA extracted from various species of squamates using a radiolabelled *28S* rRNA probe.

Supplementary Table 1. Summary of voucher numbers, number of *SPIN* sequences sequenced, distances between *SPIN* copies, geographic distribution at the time of *SPIN* HTs and *SPIN* copy numbers for all species included in this study. The literature used to infer the geographic distribution at the time of the *SPIN* transfers (within the last 50 MY) is cited in material and methods.

Supplementary Table 2. Interfamilial distances between *SPIN* consensus, *RAG1* and *C-mos* sequences for all possible pairwise comparisons of squamate families where *SPIN* was detected.

Supplementary Dataset 1. Alignment of *C-mos* sequences used to calculate pairwise distances between the various squamate families in which *SPIN* was detected.

Supplementary Dataset 2. Alignment of *RAG1* sequences used to calculate pairwise distances between the various squamate families in which *SPIN* was detected.

Supplementary Dataset 3. Alignment of *SPIN* autonomous consensus sequences used to calculate the pairwise distances between the various squamate families in which *SPIN* was detected.

Supplementary Dataset 4. Alignment of *SPIN* autonomous individual sequences used to calculate pairwise distances between the various *SPIN* copies sequenced in this study within the various squamate species.

Supplementary Dataset 5. Alignment of non-autonomous squamate *SPIN* sequences produced in this study. All sequences are aligned to the *SPIN* superconsensus reconstructed in Pace et al (2008). The internal regions of each non-autonomous element do not derive from their cognate autonomous *SPIN*, they are species specific and therefore unalignable to the *SPIN* superconsensus.

Supplementary Dataset 6. Alignment of individual autonomous *SPIN* sequences used to reconstruct the phylogeny presented in Figure 4.