The Evolution and Diversity of DNA Transposons in the Genome of the Lizard *Anolis carolinensis*

Peter A. Novick[1,2], Jeremy D. Smith[3], Mark Floumanhaft[1], David A. Ray[3], and Stéphane Boissinot*[,1,2]

[1]Department of Biology, Queens College, the City University of New York

[2]Graduate School and University Center, the City University of New York

[3]Department of Biochemistry and Molecular Biology, Mississippi State University

*Corresponding author: E-mail: stephane.boissinot@qc.cuny.edu.

Accepted: 26 November 2010

Abstract

DNA transposons have considerably affected the size and structure of eukaryotic genomes and have been an important source of evolutionary novelties. In vertebrates, DNA transposons are discontinuously distributed due to the frequent extinction and recolonization of these genomes by active elements. We performed a detailed analysis of the DNA transposons in the genome of the lizard *Anolis carolinensis*, the first non-avian reptile to have its genome sequenced. Elements belonging to six of the previously recognized superfamilies of elements (*hAT*, *Tc1/Mariner*, *Helitron*, *PIF/Harbinger*, *Polinton/Maverick*, and *Chapaev*) were identified. However, only four (*hAT*, *Tc1/Mariner*, *Helitron*, and *Chapaev*) of these superfamilies have successfully amplified in the anole genome, producing 67 distinct families. The majority (57/67) are nonautonomous and demonstrate an extraordinary diversity of structure, resulting from frequent interelement recombination and incorporation of extraneous DNA sequences. The age distribution of transposon families differs among superfamilies and reveals different dynamics of amplification. *Chapaev* is the only superfamily to be extinct and is represented only by old copies. The *hAT*, *Tc1/Mariner*, and *Helitron* superfamilies show different pattern of amplification, yet they are predominantly represented by young families, whereas divergent families are exceedingly rare. Although it is likely that some elements, in particular long ones, are subjected to purifying selection and do not reach fixation, the majority of families are neutral and accumulate in the anole genome in large numbers. We propose that the scarcity of old copies in the anole genome results from the rapid decay of elements, caused by a high rate of DNA loss.

Key words: transposon, transposase, recombination, Anolis.

Introduction

Transposable elements (TEs) are mobile DNA sequences that are typically classified into two categories: class 1 elements that require an RNA intermediate for their transposition (e.g., retrotransposons) and class 2 elements that transpose as single or double-strand DNA (Craig et al. 2002). Class 2 transposons can be divided into three categories that differ in their mode of transposition: the classical cut-and-paste DNA transposons, the rolling circle transposons or *Helitrons* and the "self-synthesizing" *Mavericks* or *Polintons* (Kapitonov and Jurka 2005; Feschotte and Pritham 2007). The cut-and-paste group is the most diverse and is further divided into ten superfamilies that diverged before the diversification of eukaryotes. Cut-and-paste transposons contain an open reading frame (ORF) that encode for the enzyme transposase. The transposase specifically recognizes the terminal inverted repeats (TIRs) of the element, excises the transposon, and inserts it elsewhere in the host's genome (Robertson 2002). Upon insertion in the genome, target site duplications (TSDs) are produced. The length and sequence of the TSDs and terminal motifs of the TIRs are highly conserved across superfamilies and are useful in categorizing elements. Following the excision of an element, the donor site may be repaired via homologous recombination. However, the gap repair process is oftentimes interrupted resulting in shorter elements with internal deletions (Engels et al. 1990). These shorter copies still possess TIRs that can be recognized by the transposase encoded by complete elements and consequently they retained their mobility (Hartl et al. 1992). These nonautonomous elements compete with their progenitors for the transposase and often outnumber their autonomous relatives (Yang et al. 2009).

TEs have dramatically affected the size, structure, and function of the genomes they inhabit (Feschotte and Pritham 2007; Cordeaux and Batzer 2009). Although most TE insertions are either neutral or deleterious, the domestication by the host of TE-encoded sequences can occur and is responsible for the evolution of fundamental biological processes such as light sensing in plants (Hudson et al. 2003; Lin et al. 2007) and V(D)J recombination in vertebrates (Jones and Gellert 2004; Mathews 2006; Jiao et al. 2007). However, it is likely that the impact of class 1 and class 2 elements varies among species because their abundance and diversity greatly differ among group of organisms (Eickbush and Furano 2002; Furano et al. 2004; Pritham et al. 2005). For instance, fish genomes contain a diversity of active DNA transposons that coexist with a multitude of retrotransposon families (Duvernell et al. 2004). In contrast, mammalian genomes are dominated by class 1 elements and it was believed until recently that mammals completely lack active class 2 elements, although DNA transposons were once diverse and very active in early mammalian evolution (Lander et al. 2001; Lindblad-Toh et al. 2005; Pace and Feschotte 2007). However, recent analyses have shown that vertebrate genomes, including mammalian genomes, can be recolonized by laterally transferred DNA transposons and that these transfers seem to occur relatively frequently (Pace et al. 2008; Gilbert et al. 2010; Novick et al. 2010).

Here, we present the first analysis of class 2 elements in a reptile, the North American green anole, *Anolis carolinensis*. The green anole is the first non-avian reptile to have its genome sequenced, bridging a large phylogenetic gap between fish and mammals. We discovered that DNA transposons are represented in the anole genome by six superfamilies (*hAT*, *Tc1/Mariner*, *Helitron*, *PIF/Harbinger*, *Polinton/Maverick*, and *Chapaev*). However, only four of them have amplified to significant numbers (*hAT*, *Tc1/Mariner*, *Helitron*, and *Chapaev*). These prolific superfamilies are represented in the anole genome by ten autonomous families, which are responsible for the amplification of a multitude of nonautonomous families that largely outnumber their autonomous counterparts. The age distribution of DNA transposons suggests that novel insertions do reach fixation, yet the near absence of ancient elements indicates that some postinsertional mechanism(s) limits the accumulation of DNA transposons in the anole genome.

Materials and Methods

Acquisition of Class 2 Elements in the Anole Genome

An exhaustive search of the anole genome for class 2 transposons was completed with three different methods. Our initial analysis was accomplished using the program PILER (Edgar and Myers 2005). We used this program to find matching sequences of a minimum repeat length of 100 bp and a minimum consensus of 95%. From the output subgroups, only those with a minimum of ten copies were analyzed. These novel families were then extracted and assembled into contigs using Seqman II (part of the DNAStar package: http://www.dnastar.com/), combining any redundant output into a single family. The resulting alignments were collected to form an initial library of TEs from the anole genome. This library was then used as the basis of a RepeatMasker search of the genome to find additional copies of the putative elements. Hits of sufficient length, typically at least 100 bp, were extracted from the genome along with a minimum of 500 bp of flanking sequence using custom PERL scripts. The extracted sequence subgroups were again aligned using MUSCLE (Edgar 2004) and consensus sequences were generated. The process was repeated until the full-length sequence of each putative element was obtained.

A second search of the genome, using the Repeatscout program (Price et al. 2005), was performed to identify any previously undiscovered elements (lmer = 12). The resulting putative TEs were assembled into contigs using Seqman II and aligned using MUSCLE. Any previously identified putative elements were not processed, whereas new elements were used to create a library for use in a Repeatmasker search and the output as described above.

Lastly, we performed a BlastX search of the genome using amino acid sequences derived from a known transposon library available from Repbase (v13.01). The resulting hits of at least 100 bp, with a maximum score of 2×10^{-150}, at multiple loci were extracted along with 500 bp of flanking sequence (using a modified version of the previously used PERL scripts) and aligned.

Classification of Elements

Elements were separated into superfamilies and further subdivided into families based upon size and sequence similarity. A consensus sequence for each family was created. The pairwise divergence between elements and the average divergence from the consensus sequence were calculated using Kimura's 2-parameter method in MEGA 4.0 (Tamura et al. 2007). We estimated copy number for each family by using the Blast option on NCBI; however, as many of the elements are either extremely fragmented and some families are nearly identical at their ends but differ in their central region, it is difficult to ascertain the exact copy number of each family. In order to only count nonfragmented elements, we limited our percent identity score to 90%.

Finally, elements were scanned for the presence/absence of TIRs, TSDs, ORFs, and similarity to elements in other genomes. TIRs were discovered by aligning the 5′ in the positive orientation with the reverse complement of the 3′ of elements belonging to the same family. TSDs were determined by collecting 20 bp downstream of the 3′ end and 20 bases upstream of the 5′ end for at least 20 sequences

per family. Percentages for each of the four possible nucleotides were then calculated for each position. As repeat masker oftentimes did not recognize some of these novel nonautonomous elements, the sequence and length of TSDs and TIRs were used to categorize each family to its proper autonomous superfamily. ORF finder and Conserved Domains Database (Marchler-Bauer et al. 2007) were used in tandem to find the length of ORFs and types of proteins encoded by autonomous elements. Finally, in order to identify possible events of horizontal transfer, consensus sequences from each family were submitted to the Blast option on the NCBI Web site and a multitude of sequenced organisms were screened.

Results

The genome of *A. carolinensis* contains elements representative of six superfamilies, but two of these superfamilies (*PIF/Harbinger* and *Polinton/Maverick*) did not produce any significant amplifications and are represented by less than 10 copies. Conversely, the *hAT*, *Mariner*, *Helitron*, and *Chapaev* superfamilies were very prolific and produced 67 distinct families (ten autonomous and 57 nonautonomous), yet they differ drastically in abundance and diversity.

The *hAT* Superfamily

The *hAT* superfamily is the most abundant and diverse in the anole genome. It is represented by five autonomous and 32 nonautonomous families (table 1). All these families display the structural features typical of the *hAT* superfamily, including 8 bp TSDs and terminal motifs of YARNG. Four of the five autonomous *hAT* families are found in distantly related animals and result from independent events of horizontal transfer (*hAT-HT1_AC*, *hAT-HT2_AC*, *hAT-HT3_AC*, and *SPIN_AC*) (Pace et al. 2008; Novick et al. 2010). These four laterally transferred *hAT* elements are more closely related to mammalian *Charlie* elements (fig. 1). The fifth family, called *hobo_AC*, is the only one for which there is no evidence of lateral transfer because we failed to find similar elements in other genomes. As its name indicates, it belongs to the *hobo* clade of *hATs* (fig. 1). Each autonomous *hAT* family produced nonautonomous (from 1 to 15) families with similar TSD and TIR sequences. Nonautonomous copies outnumber autonomous copies in the anole genome by nearly two orders of magnitude (~330 autonomous copies vs. ~24 000 nonautonomous copies). For instance, the most abundant autonomous family, *hobo_AC* (290 copies), is responsible for the mobility of at least 15 nonautonomous families, totaling 3,272 copies. Most nonautonomous families correspond to deleted versions of autonomous elements; consequently their evolutionary affinities are relatively easy to determine. Yet, we identified eight nonautonomous families that did not show any similarity with a known autonomous family beyond the TIR. For instance, *hAT-N4_AC* elements have TIRs that are indistinguishable in length and sequence from *hAT-HT2_AC*, yet they do not share homology with autonomous and nonautonomous *hAT-2* outside the TIRs (table 1). Although the origin of this family remains unclear, the similarity of its TIRs with the TIRs of autonomous *hAT-HT2* elements suggests that *hAT-N4_AC* elements are mobilized by *hAT-HT2*. Of the eight "orphan" families, we found similarity with the TIR of an autonomous family for five of them suggesting that a known autonomous family is responsible for their mobilization. The remaining three (*hAT-N6*, *hAT-N7*, and *hAT-N8*) have identical TIRs, but the sequence of their TIR is different from the TIR of the five autonomous families. The transposase of one of the autonomous copies could be mobilizing these elements despite their lack of similarity in the TIR. Alternatively, we cannot exclude that another autonomous *hAT* family exists in anole and either was missed by our search if its copy number is extremely low or is so new in anole that it is polymorphic in populations and absent from the individual used for the genome sequencing or has never reached fixation in the more distant past.

Although autonomous *hAT* elements in anole are typical members of their superfamily (length and sequence of TIRs and TSDs), they differ considerably in length. The *hAT-HT1_AC*, *hAT-HT2_AC*, *hAT-HT3_AC*, and *SPIN_AC* families are all between 2 and 3 kb long, but *hobo_AC* elements are much longer, ranging from 9 to 15 kb (fig. 2). This unusual length results from the incorporation in *hobo* elements of a considerable amount of extraneous DNA, including a number of partial transposon insertions from class 2 (such as multiple fragments of *Chapaev3-1_AC*) and class 1 (such as *RTE Bov-B*, *CR1*, and *Sauria-SINE*) (Kordis and Gubensek 1998; Piskurek et al. 2006; Shedlock 2006). As these elements belong to families older than *hobo_AC*, it is likely that their presence in the sequence of *hobo_AC* elements results from the incorporation of large fragments of genomic DNA that fortuitously contained ancient TE fragments and not from insertion events in *hobo_AC* elements. The incorporation of genomic DNA has drastically increased the length of *hobo_AC*, yet it has not altered the replicative ability of this family. The phylogenetic tree in figure 2 recapitulates the evolution of the *hobo_AC* family. It is based on a 3-kb fragment common to all full-length *hobo_AC* elements that includes the transposase domain. Autonomous *hobo* elements are very similar in this 3 kb region, as suggested by the short length of the branches, yet they cluster in several distinct lineages that differ significantly in structure due to the frequent insertion or loss of DNA sequence (fig. 2). The extreme structural variation in the relatively young *hobo_AC* family demonstrates that DNA transposons in anole can gain or lose DNA sequences at a very high rate. Additionally, we found that some elements were composites of other autonomous copies. For instance, the 5′ half of element 1 (fig. 2) is very similar to element 3 but its 3′ half is identical in structure to element 4. This indicates that

Table 1
Characteristics and Nomenclature of all Families of Autonomous and Nonautonomouos *hAT/hobo* DNA Transposons in the Lizard *Anolis carolinensis*.

Name	Copy Number >90% Identity	Length in bp	TSD	Length of TIR	TIR	% Divergence ± SE	% Divergence from Consensus ± SE
hAT-HT1_AC	5 (5)[a]	2968	8 bp[b]	16	CAGTGATGGSSAACCT	5.82 ± 0.40	3.88 ± 0.35
hAT-HT1N1_AC	868	585	NTCTAGAN	16	CARTGATGGSCAACCT	4.75 ± 0.64	2.55 ± 0.42
hAT-HT2_AC	5 (4)[a]	2246	8 bp[b]	15	CAGGGGTCCCCAAAC	0.14 ± 0.07	0.00 ± 0.00
hAT-HT2N1_AC	28	1485	NHCTAGRN	16	CAGGGGTCCCCAAACT	1.34 ± 0.31	0.73 ± 0.15
hAT-HT2N2_AC	43	1065	NTCTAGAN	16	CAGGGGTCCYCAAACT	3.09 ± 0.27	1.52 ± 0.22
hAT-HT2N3_AC	1,372	780	NTNTANAN	16	CAGGGGTCCYCAAACT	4.76 ± 0.43	1.75 ± 0.18
hAT-HT3_AC	<25 (13)[a]	2754	8 bp[b]	14	CAGTGRTTCCCAAA	5.62 ± 0.29	3.22 ± 0.22
hAT-HT3N1_AC	344	326	NTCTAGAN	14	CAGTGRTTCCCAAA	3.98 ± 0.37	1.78 ± 0.23
hAT-HT3N2_AC	132	833	NYYTARRN	16	CAGGGGTCCCCAAACT	8.44 ± 0.77	4.04 ± 0.35
hAT-HT3N3_AC	682	799	NTCTAGAN	14	CAGTGRTTCCCAAA	9.66 ± 0.54	6.31 ± 0.42
hobo_AC	25 (290)[a]	≈15,000	8 bp[b]	11	TAGGCTTGMTC	1.93 ± 0.09	1.12 ± 0.07
hobo-N1_AC	300	1854	NTRNNYAN	11	TAGGCTTSATC	1.80 ± 0.30	0.90 ± 0.15
hobo-N2_AC	201	2359	NTRNNYAN	11	TAGGCTTSATC	1.71 ± 0.40	1.08 ± 0.30
hobo-N3_AC	489	2192	NTRNNYAN	11	TAGGCTTSATC	0.90 ± 0.01	0.50 ± 0.01
hobo-N4_AC	60	2042	NTRNNYAN	11	TAGGCTTSATC	1.07 ± 0.09	0.49 ± 0.06
hobo-N5_AC	34	2440	NTRNNYAN	11	TAGGCTTSATC	0.88 ± 0.09	0.27 ± 0.04
hobo-N6_AC	188	1995	NTRNNYAN	11	TAGGCTTSATC	0.82 ± 0.11	0.28 ± 0.14
hobo-N7_AC	295	2766	NTRNNYAN	11	TAGGCTTGAGC	1.64 ± 0.13	0.46 ± 0.04
hobo-N8_AC	264	2262	NTRNNYAN	11	TAGGCTTGAGC	2.28 ± 0.21	0.73 ± 0.08
hobo-N9_AC	284	2465	NTRNNYAN	11	TAGGCTTGAGC	1.90 ± 0.21	1.04 ± 0.13
hobo-N10_AC	71	2646	NTRNNYAN	11	TAGGCTTGAGC	2.43 ± 0.18	0.69 ± 0.11
hobo-N11_AC	120	2686	NTRNNYAN	11	TAGGCTTGAGC	1.92 ± 0.13	0.68 ± 0.19
hobo-N12_AC	26	3413	NTRNNYAN	11	TAGGCTTGAGC	1.81 ± 0.13	0.91 ± 0.08
hobo-N13_AC[c]	16	>3.5 kb	NA	11	TAGGCTTGAGC	2.23 ± 0.23	1.15 ± 0.15
hobo-N14_AC	235	1911	NTANNTAN	11	TAGGCTTGMKC	1.06 ± 0.14	0.56 ± 0.08
hobo-N15_AC	689	2490	NTANNTAN	11	TAGGCTTGAKC	0.56 ± 0.10	0.23 ± 0.04
SPIN_AC[b,d]	<5 (1)[a]	NA	—	—	CAGYGGTTCTCAACCT	NA	
SPIN_NA11_AC[b,d]	12,138	273	—	—	—	NA	NA
SPIN_NA1_AC[d]	181	745	NYYTARRN	16	CAGTGKTTCTCAACCT	12.3 ± 0.69	7.04 ± 0.74
hAT-N1_AC	>1,000	135	NYYTARRN	57	CAGTGGTTCTCAACCTGTGG[e]	4.51 ± 0.66	2.36 ± 0.37
hAT-N2_AC	<20	188	NYYTARRN	15	CAGSYTTYTYMAMCM	49.9 ± 4.49	18.5 ± 2.26
hAT-N3_AC	163	600	NYYYRRRN	16	CAGTGGTTCCCAACCT	10.8 ± 0.77	6.00 ± 0.51
hAT-N4_AC	>1,000	732	NYYTARRN	16	CAGGGGTCCYCAAACT	3.01 ± 0.24	1.50 ± 0.12
hAT-N5_AC	16	495	NTNTANAN	16	CAGGSRTGTCCAACCT	33.7 ± 2.97	15.2 ± 1.93
hAT-N6_AC	964	810	NTTRYAAN	12	CAGAGSCGGYCC	0.60 ± 0.09	0.30 ± 0.04
hAT-N7_AC	>1,000	288	NNTNNANN	12	CAGAGCCGGYCC	2.39 ± 0.36	1.68 ± 0.51
hAT-N8_AC	>1,000	329	NNTNNANN	12	CAGAGSCGGYCC	1.77 ± 0.35	1.25 ± 0.27

SE, standard error.
[a] Number of ORFs found in the genome regardless of the 5′ and 3′ ends.
[b] Not enough elements were retreived to construct a TSD pattern beyond the number of base pairs.
[c] Due to the lack of complete elements, TIR was deduced from the 3′ end only.
[d] Data from Pace et al. (2008).
[e] Only the first 20 bp of the TIR are presented here.

interelement recombination can generate novel elements, thus increasing the structural diversity of the *hobo_AC* family.

The dynamic nature of *hobo_AC* evolution is also apparent in nonautonomous copies. *hobo_AC* is responsible for the amplification of at least 15 families of nonautonomous elements ranging in length from 1.3 to 3.4 kb. These nonautonomous families can be separated into two groups that contain slightly different TIRs: group A that contains families *hobo-N1_AC* to *hobo-N6_AC* and group B with families *hobo-N7_AC* to *hobo-N13_AC* (fig. 3). The similarity between the TIRs of group A with the TIRs of autonomous *hobo_AC* elements suggests that these elements result from deletions of the autonomous elements found in the anole genome. In contrast, group B elements are likely to have evolved from a subset of *hobo_AC* elements with different TIRs that is apparently no longer present in the anole genome. Within group A and B, elements have similar ends but differ drastically in structure due to insertions, deletions, and the incorporation of genomic DNA of other origin, often

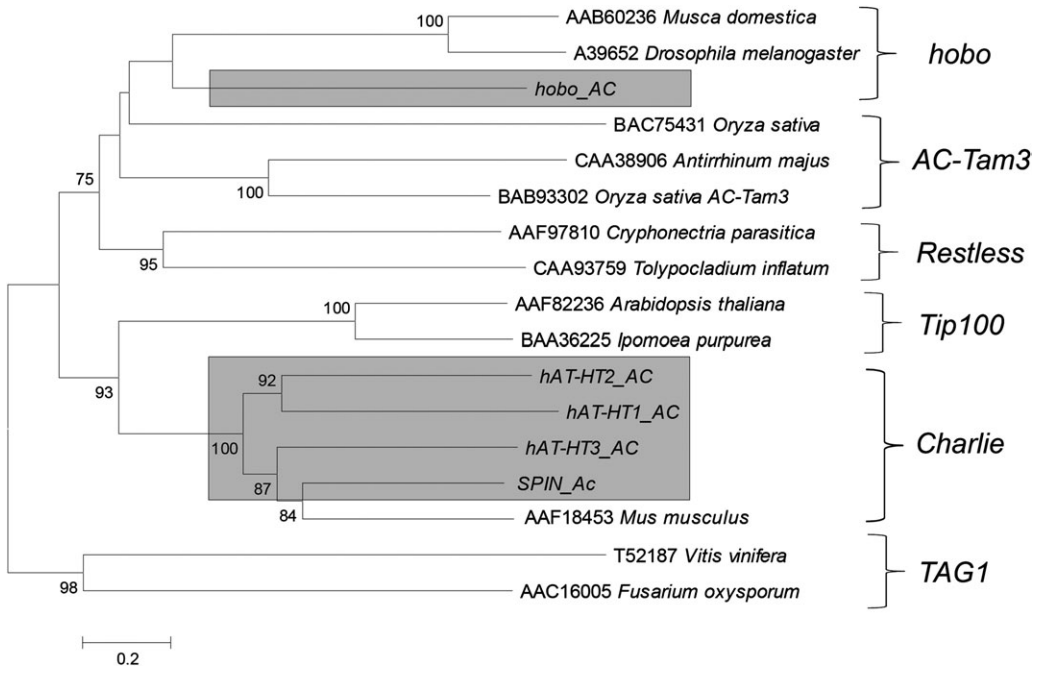

Fig. 1.—Phylogenetic position of *Anolis hATs* relative to previously described *hAT* families. The tree is based on an amino acid alignment of the transposase domain. It was inferred using the neighbor joining method, and the robustness of the nodes was assessed by bootstrap (1,000 runs). Bootstrap values <75% have been removed.

containing TE from other classes or superfamilies. Using the presence of TE fragments embedded within nonautonomous families as markers, we were able to decipher the evolutionary history of these families (depicted on fig. 4). The ancestral group A elements contained a 520 bp *Penelope* insertion (Arkhipova 2006) and a 333 bp *ACA SINE*. A partial deletion (250 bp) of the *Penelope* element occurred yielding families *hobo-N2*, *3*, and *5*. Independently, a partial deletion (176 bp) of the *ACA SINE* occurred and is shared by families *hobo-N1* and *6*. A recombination event between an element containing the deleted version of *Penelope* and one containing the deleted *SINE* resulted in family *hobo-N4*. Ancestral group B elements contain a *CR1* and a *Chapaev* insertion and are represented in the anole genome by families *hobo-N7*, *8*, and *9*. Two independent recombination events occurred between group B and group A elements: one resulted in a family that is identical to a typical group A element but with a 3′ end similar to a group B element (*hobo-N15_AC*) and the other produced a family which is similar to group B over most of its length but has a group A 3′ end (figs. 3 and 4). A fourth recombination event between family *hobo-N15* and a group B element produced three families (*hobo-N10*, *11*, and *12*) with termini typical of group B but a central portion similar to group A.

The vast majority of *hAT* elements are young, as the average divergence from consensus is less than 10% for 85% of the families (fig. 5). As expected, there is a relatively good concordance between the age of autonomous families and their nonautonomous counterparts. This is true of families *hAT-HT1_AC*, *hAT-HT3_AC*, and *hobo_AC*. For instance, the average divergence from consensus of family *hobo_AC* is 1.12%, whereas the divergence of its nonautonomous relatives ranges from 0.49% to 1.15%. However, this might not be true of family *hAT-HT2_AC*. Autonomous *hAT-HT2* elements are extremely young and in fact their mean divergence from consensus is 0.00%. In contrast, their nonautonomous counterparts have divergence ranging from 0.73% to 1.75% and, thus, seem to predate their autonomous progenitor. It is plausible that these nonautonomous copies resulted from a previous wave of lateral transfer of the *hAT-HT2* family that would have produced nonautonomous families but failed to establish a resident population of autonomous copies. Alternatively, this suggests that autonomous and nonautonomous copies have different dynamics in *Anolis* populations, possibly because they are differently affected by purifying selection. If autonomous copies are more deleterious to the host than nonautonomous ones, it is plausible that they fail to reach fixation and that their very young age reflects a high rate of turnover where the transposition of new copies is countered by the selective loss of deleterious copies. Neutral or nearly neutral elements, such as nonautonomous ones could reach fixation and accumulate more readily in the genome of the host.

The *Tc1/mariner* Superfamily

The second most diverse superfamily of DNA transposons, the *Tc1/mariner* superfamily, consists of one autonomous family and 17 nonautonomous families. These families are

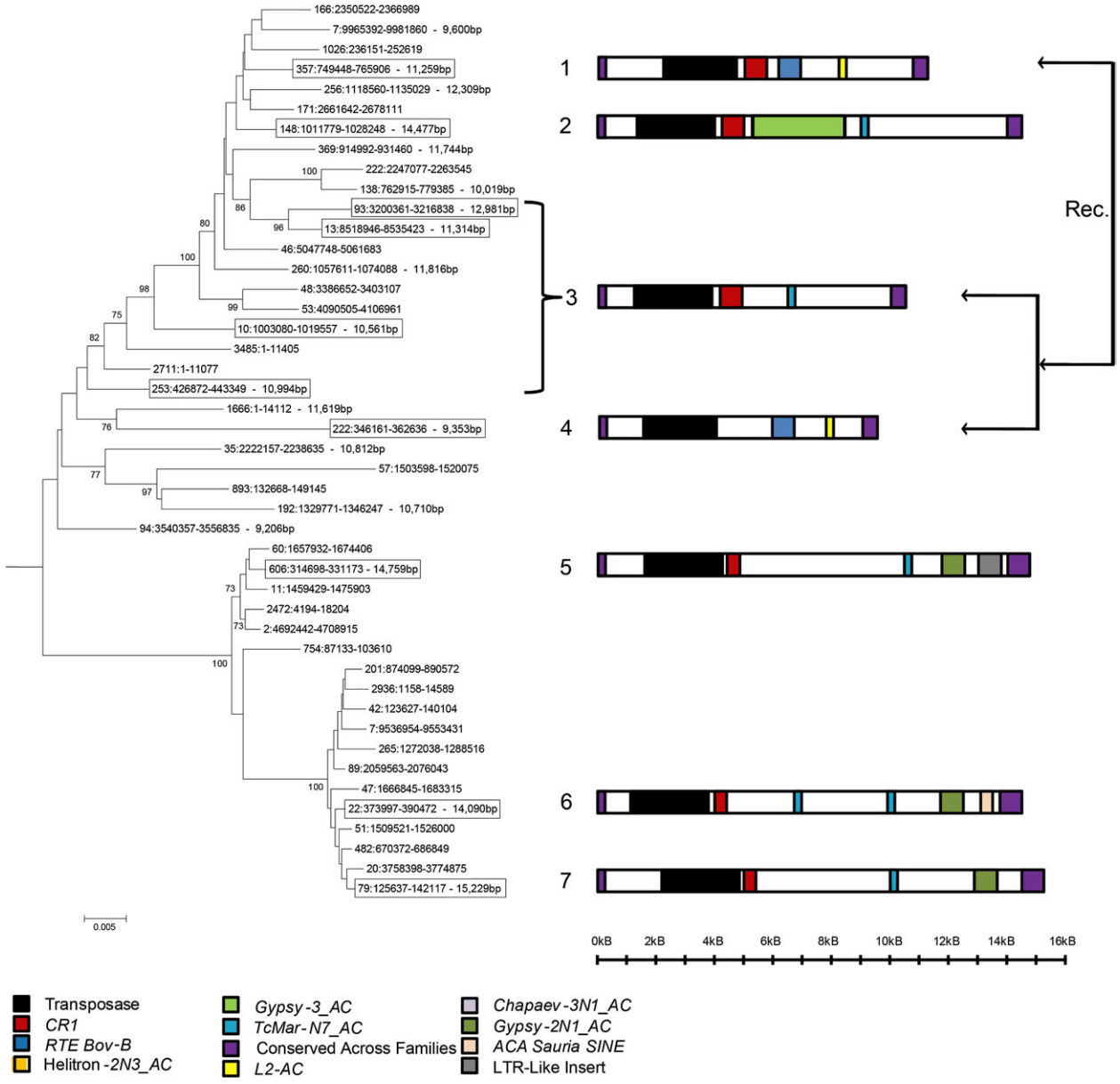

Fig. 2.—Neighbor joining phylogeny of autonomous *hobo_AC* elements based on 3 kb of the transposase domain. The boxed sequences indicate elements that are complete in the genome assembly we used. The presence of nested TE in complete *hobo_AC* elements was determined by running Repeatmasker with a library of repetitive sequences found in the anole genome. Seven different patterns of nested elements were recovered and are schematically presented on the right of each sequence (structure 3 corresponds to elements 93, 13, 10, and 253). Though all 45 elements are very similar to each other, they differ in their length and structure. The arrows on the right indicate the recombination of elements 3 and 4 resulting in element 1.

characterized by a TA TSD and TIRs that vary considerably in length and sequence. The single autonomous family, *Tc1-1_AC*, contains 58 copies and is related to the *Tc1* subset of *Mariner* elements (fig. 6). The 17 nonautonomous *Tc1*-like families range from 14 to ~5,000 copies for a total of ~17,500 copies (table 2). Surprisingly, none of the nonautonomous elements share any similarity with *Tc1-1_AC*. *Tc1-1_AC* is relatively ancient as elements diverge from each other

→

region (right) of nonautonomous *hobo_AC* elements. Bootstrap values less than 75% have been removed. At least three elements from each family are included. Boxes around elements reveal the group swap of *hobo-N14_AC* (blue) and *hobo-N15_AC* (red) in group A (light red) and group B (light blue) due to interelements recombination.

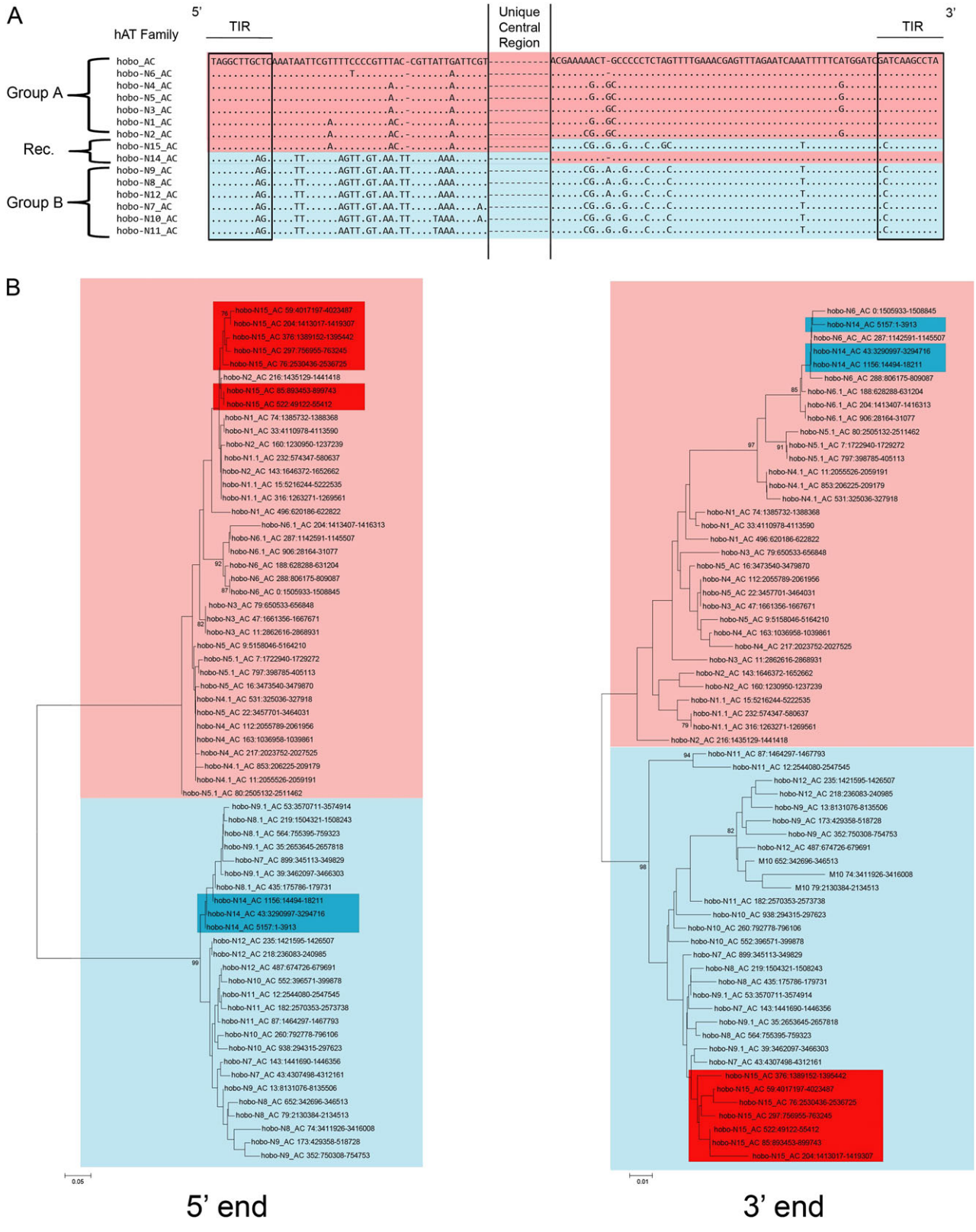

Fig. 3.—(*A*) 5′ and 3′ termini of consensus sequences of autonomous and nonautonomous *hobo_AC* families. The TIRs are boxed. Although these 16 families have similar 5′ and 3′ ends, they differ considerably in their central region due to a large number of indels and transposon insertions. Thus, the central region is unique and specific of a given family; (*B*) Neighbor joining trees based on 150 bp of the 5′ region (left) and 300 bp of the 3′

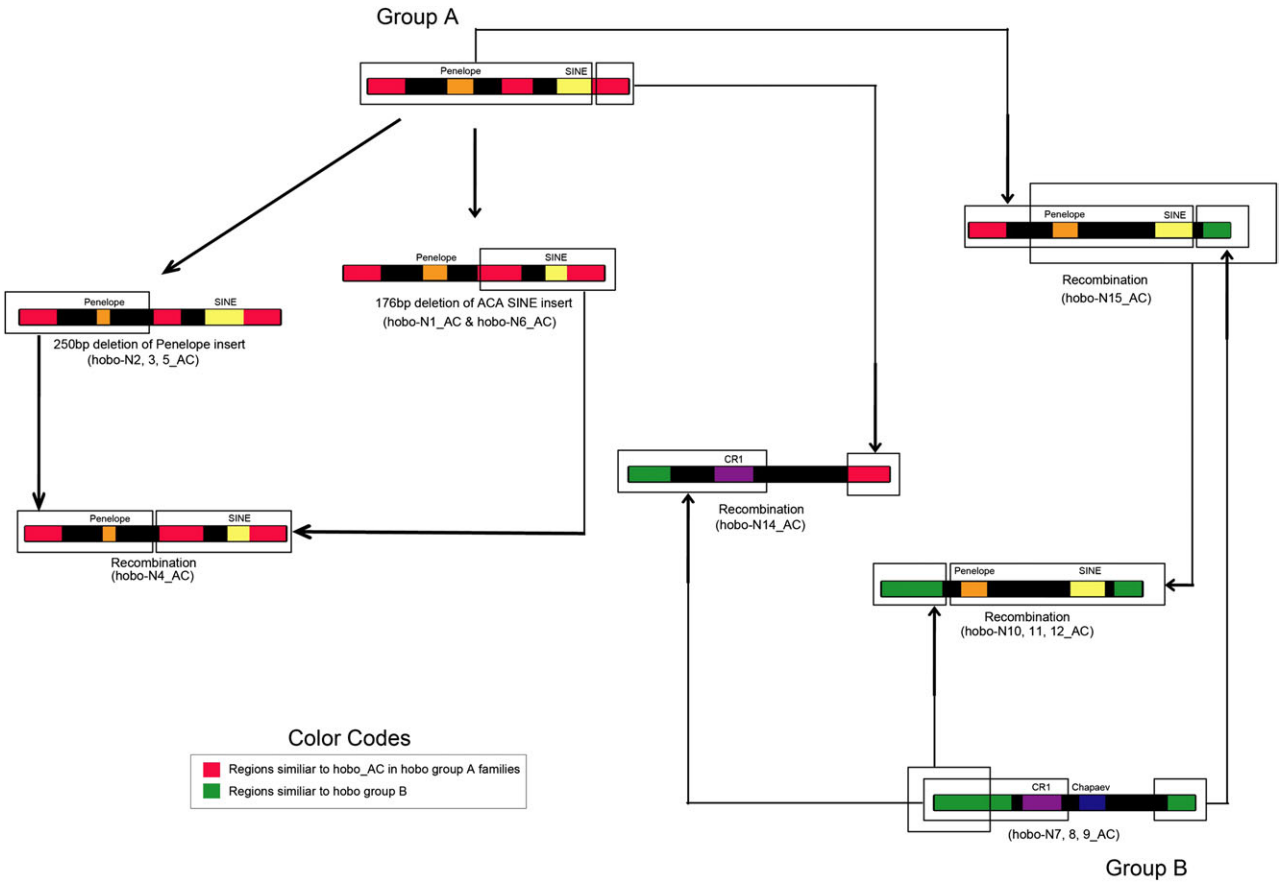

Fig. 4.—Diagram depicting the evolution of the 15 nonautonomous *hobo* families (see text for explanation).

by 13.0% on average and is probably inactive. Thus, it is unlikely that *Tc1-1_AC* is responsible for the recent burst of activity of several families (*TcMar-N3_AC*, *TcMar-N4_AC*, *TcMar-N5_AC*, *TcMar-N7_AC*, and *Tcmar-N16_AC*) that amplified to considerable numbers and have very low average divergence (less than 5%). It is also unlikely that *Tc1-1_AC* is responsible for the amplification of ancient families that predates its origin (such as *TcMar-N13*, *14*, *15*, and *17*). This suggests that other *Mariner* families are or have been active in the anole genome. Despite numerous attempts to identify such autonomous families, we failed to find any other autonomous families. However, several nonautonomous families show some significant similarity with *Mariner*-like elements from other organisms, in particular at their extremities. For instance, families *TcMar-N11* and *17* have ends similar to marsupial *Tigger3* and *Tigger4* elements, respectively, whereas *TcMar-N4* has similarity with the *Tc2*-related *DNA5_Xt* element from the frog genome. This suggests that a diversity of *Tc1/Mariner*-related elements have recently been active in the anole genome. It is plausible that these elements never reached high copy numbers and, because of the fast decay of TEs in anole (Novick et al. 2009), are no longer identifiable. Unlike *hAT* elements, most *Mariner*-like families are relatively ancient and no longer active. In fact, only five families have an average pairwise divergence lower than 5% and only two diverge by less than 3%. The rarity of autonomous copies and the age distribution of *Mariner*-like families suggest that mariner elements could be frequently invading the anole genome, produce abundant nonautonomous families but fail to become stable residents of this genome.

The *Helitron* Subclass

The third category of DNA transposon found in anole is the *Helitron* subclass (table 2). It is represented by two autonomous families, *Helitron-1_AC* and *Helitron-2_AC*, both in very low copy number (less than five copies) and six abundant nonautonomous families. All these elements are typical members of the *Helitron* superfamily, lacking TIRs and with TA TSDs (Kapitonov and Jurka 2001). Unlike *hAT* and *Mariner* nonautonomous elements, all nonautonomous *Helitron* families are internally deleted versions of autonomous families; *Helitron-1N1_AC* is derived from *Helitron-1_AC*, whereas the remaining five nonautonomous families (*Helitron-2N1-5_AC*) resulted from large deletions of *Helitron-2_AC*. As previously noted by Piskurek et al. (2009), *Helitrons* also have the ability to capture extraneous DNA.

Family *Helitron-2N3* contains fragments of *Poseidon* and *SINE* elements, yet this seems to be a rare occurrence as this is the only *Helitron* family to show a composite structure. *Helitrons* appear to have been recently active as suggested by their relatively low level of divergence and, as expected, the amplification of most nonautonomous families is concomitant to the activity of their autonomous progenitors. The only exception is the oldest family, *Helitron-2N1_AC*, which divergence lies clearly outside the divergence distribution of all other families. Yet, this family is unambiguously related to the autonomous *Helitron-2_AC* family. This suggests that the *Helitron-2_AC* family has been active in the anole genome much longer than its current divergence suggests. It is possible that *Helitron-2_AC* is in fact an ancient resident of the anole genome but remained in such low copy numbers that older copies are no longer identifiable. However, it is surprising that *Helitron-2_AC* failed to produce additional nonautonomous families after *Helitron-2N1_AC*. It is thus possible that, once a resident of the anole genome, *Helitron-2_AC* became extinct and only recently recolonized the anole genome, possibly through lateral transfer. This in fact appears to be the case. These elements are phylogenetically more closely related to insect helitrons than to other vertebrate helitrons and their high level of similarity to insect helitrons is best explained by lateral transfer from insects (Kordis 2009; Thomas et al. 2010).

The *Chapaev* Superfamily

The recently described *Chapaev* superfamily (Bao et al. 2009) is also present in the anole genome and consists of four families, two of which are autonomous. Elements in this superfamily display a 3 bp TSD of consensus TWA and TIRs beginning with the trinucleotide CAC. They belong to the subsets of elements classified as *Chapaev3* in Repbase. These families are the oldest one we found and are no longer active: the oldest family contains elements that are over 35% divergent, whereas the youngest contains elements that are no less than 12% divergent (table 2). Although these families are old, ORFs are still detectable in the two autonomous families, *Chapaev3-1_AC* (1910 bp in total length) and *Chapaev3-2_AC* (1767 bp in total length) of 485 and 561 amino acids, respectively. The two remaining families of *Chapaev* elements are nonautonomous families (*Chapaev3-3N1_AC* and *Chapaev3-3N2_AC*) and are also ancient, yet they are not directly derived from the autonomous families.

The Polinton/Maverick and PIF/Harbinger Superfamilies

As our search of the anole genome was targeted toward repetitive sequences, we missed two superfamilies, *PIF/Harbinger* and *Polinton/Maverick*, represented by very low (<10) copy numbers. These two superfamilies had previously been discovered in the lizard genome by Kordis (2009). We specifically searched for these elements and collected them.

We identified seven copies belonging to the *Polinton/Maverick* superfamily; unfortunately all but one copy are incomplete to various degrees due to gaps in the available genome sequence. The only complete element is 17,487 bp long and harbors the features typical of *Polinton* elements, including a 6 bp TSD and long (584 bp) TIR (Kapitonov and Jurka 2005). This element contains eight ORFs, which is typical of *Polintons*. The position and orientation of the ORFs are identical to the sea urchin *Polinton-1_SP* and the zebrafish *Polinton-1_DR*. We found two very incomplete elements with high similarity to this full-length copy. These three copies diverged from each other by 0.2–0.3% suggesting that they constitute a very small (<5 copies), yet active *Polinton* family in the lizard. A second near full-length element (>11 Kb of sequence available) was also recovered but it differs from the only full-length copy by 11.65%, yet it also contains eight intact ORFs intact and is thus potentially active. The fact that lizard *Polinton* have protein-coding abilities indicates that they are probably recent insertions and the presence of a very small number of closely related copies suggest that *Polintons* are indeed active in anole. However, the very low copy number of *Polintons* in anole and the apparent absence of nonautonomous copies suggest that *Polintons* are either newcomers in the anole genome or are subjected to a high rate of turnover.

The anole genome contains five ORFs (442–466 amino acids) related to transposases of another superfamily, *PIF/Harbinger*. These ORFs differ from each other by 53–68% at the amino level, yet most of the motifs diagnostic of *Harbinger*'s transposases are conserved. These copies are not flanked by the 14–25 TIRs and the TTA/TAA TSDs characteristic of the *Harbinger* family and do not correspond to active transposon copies. Because they have retained their protein-coding ability, it is very likely that these ORFs represent instances of domestication of now-extinct *Harbinger* transposons. A single case of *Harbinger* transposase domestication had previously been reported in vertebrates (Kapitonov and Jurka 2004) and the five new cases reported here emphasize the potential of *Harbinger* transposase as a source of protein motifs that can be recruited by the host.

Discussion

The genome of *A. carolinensis* harbors an extraordinary diversity of active or recently active class 2 transposons. Six superfamilies (*hAT*, *Tc1/Mariner*, *Helitron*, *PIF/Harbinger*, *Polinton/Maverick*, and *Chapaev*) are represented in the anole genome and four of them (*hAT*, *Tc1/Mariner*, *Helitron*, and *Chapaev*) have amplified to considerable numbers, generating a total of 67 families, including 10 autonomous ones. Two superfamilies are extinct in anole: the *Chapaev* and the *PIF/Harbinger* superfamilies. *Chapaev* has produced

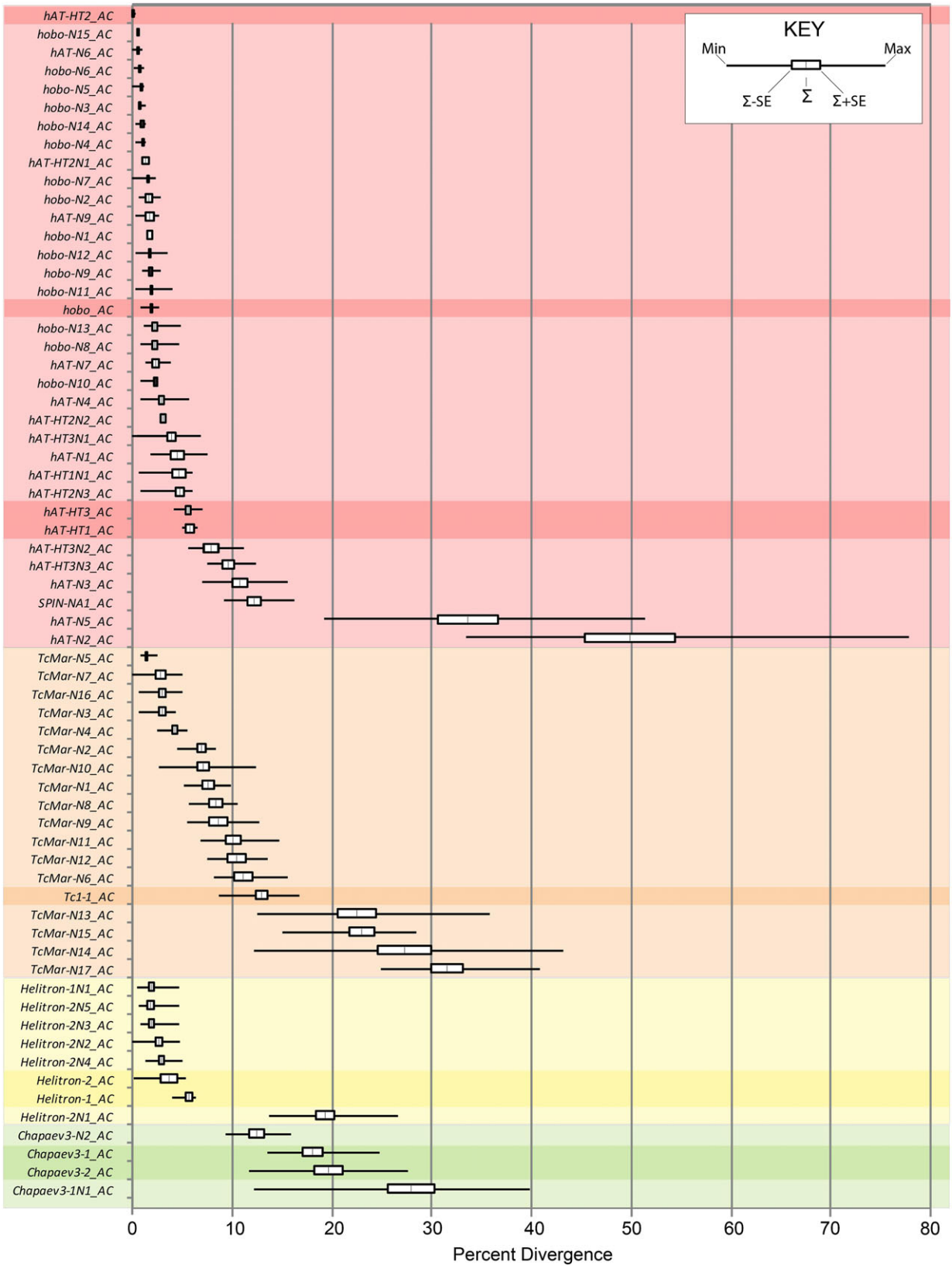

Fig. 5.—Divergence plot of *hAT* (red), *Tc1/Mariner* (orange), *Helitron* (yellow), and *Chapaev* (green) families found in the genome of the lizard. Values were calculated using Kimura's 2-parameter method in Mega 4.0. Autonomous families are emphasized with darker bars.

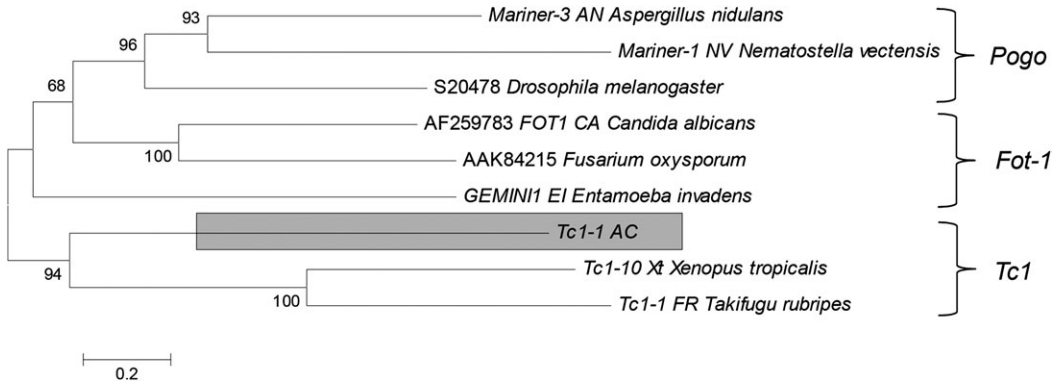

Fig. 6.—Phylogenetic position of the only autonomous *Tc1/Mariner* family in anole relative to previously described *Mariner* families. The tree is based on an amino acid alignment of the transposase domain. It was built using the neighbor joining method and the robustness of the nodes was assessed by bootstrap (1,000 runs). Bootstrap values <65% have been removed.

several distinct autonomous and nonautonomous families. In contrast, *PIF/Harbinger* is represented by only five domesticated and very divergent copies. Among the four superfamilies that have recently been active, one of them, *Polinton*, is represented by a very small number of copies, whereas the other three (*hAT*, *TC1/Mariner*, and *Helitron*) have been extremely prolific, increasing the genetic diversity of the green anole. The *hAT*, *Tc1/Mariner*, *Chapaev*, and *Helitron* superfamilies are also represented by a large diversity of nonautonomous families that are either internally deleted versions of autonomous copies (most *hAT* and all *Helitron* families) or do not show similarity with autonomous copies (some *hAT* and all *Tc1/Mariner* families). As in plants and other animals, nonautonomous families greatly outnumber their autonomous counterparts (Hartl et al. 1992; Le Rouzic and Capy 2006; Yang et al. 2009). The most active and diverse superfamily, *hAT*, exemplifies a number of mechanisms that can increase the diversity of class 2 transposons. First, the frequent lateral transfer of active *hAT* families has a dramatic impact on transposons diversity. Thirteen of 37 *hAT* families (*hAT-1*, *2*, *3* and their nonautonomous relatives) are the direct or indirect result of lateral transfer (Pace et al. 2008; Novick et al. 2010; Gilbert et al. 2010). Although we failed to find elements similar to *hobo_AC* in other genomes, we cannot exclude that this family was also laterally transferred. In fact, the recent burst of activity of *hobo* elements and the absence of old (>5% divergence) autonomous or nonautonomous *hobo* related families suggests that *hobo* is a new resident of the anole genome; hence, it is plausible that all *hATs* in anole result from lateral transfer. Lateral transfer is not limited to the *hAT* superfamily. It seems to also occur in the *Helitrons* superfamily and the transfer of *Helitrons* from insects to lizard was recently demonstrated (Kordis 2009; Thomas et al. 2010). Second, nonautonomous and autonomous *hobo* elements show considerable structural variation resulting from their ability to capture and incorporate extraneous DNA sequences. This mechanism produced autonomous elements that are the longest reported in the *hAT* superfamily and some nonautonomous *hobo* elements are longer than autonomous copies of other families. As the filler DNA often contains TEs of other classes and superfamilies, the amplification of *hobo* families significantly contributes to increasing the copy number of the elements they mobilize. The unusual ability of *hobo* to incorporate very large cargo suggests that *hobo* would be an excellent candidate for the development of DNA delivery vectors. Finally, interelement recombination seems to occur readily and is responsible for the diversification of both autonomous and nonautonomous families. In particular, recombination between nonautonomous copies resulted in the formation of six novel families of elements. The exchange of sequence between superfamilies produces structural variants that can eventually compete with each other and evolve as separate entities. This exchange of genetic information between TEs could eventually yield novel autonomous families and provides a mechanism that could account for the apparent modularity of TE's evolution (Lerat et al. 1999).

The age distribution of class 2 families differs among superfamilies and reveals different dynamics of amplification in the anole genome. The *hAT* superfamily shows the highest level of recent activity as most *hAT* elements have very low level of divergence from their consensus (<5%). In fact, some of these families have such low divergence (<0.5%) that they are certainly active in extant anoles. The lack of older *hAT*-related families is likely to reflect the fact that this genome was recently colonized by *hAT* elements through lateral transfer. It is, however, surprising that four (or maybe five) *hAT* families invaded the anole genome recently but that lateral transfer did not occur in the more distant past. The presence of a couple of ancient orphan families (>10% from consensus) suggests that episodes of lateral transfer indeed occurred but had limited success at invading the genome of ancestral anoles. The

Table 2
Characteristics and Nomenclature of DNA Transposons in the Lizard *Anolis carolinensis*.

Superfamily	Name	Copy Number >90% Identity	Length in base pair	TSD	Length of TIR	TIR	% Divergence ± SE	% Divergence from Consensus ± SE
Helitron	*Helitron-1_AC*	<5	8,771	TA	—	—	5.72 ± 0.35	3.04 ± 0.19
	Helitron-1N1_AC	84	1,436	TA	—	—	1.90 ± 0.25	1.02 ± 0.14
	Helitron-2_AC	<5	4,982	TA	—	—	3.66 ± 0.87	1.95 ± 0.40
	Helitron-2N1_AC	66	1,290	TA	—	—	19.4 ± 0.96	11.3 ± 0.66
	Helitron-2N2_AC	608	1,930	TA	—	—	2.69 ± 0.34	1.58 ± 0.27
	Helitron-2N3_AC	>1,000	2,000	TA	—	—	1.98 ± 0.29	0.85 ± 0.10
	Helitron-2N4_AC	860	550	TA	—	—	2.96 ± 0.32	2.55 ± 0.30
	Helitron-2N5_AC	>1,000	2,000	TA	—	—	1.91 ± 0.36	1.78 ± 0.35
TcMar	*Tc-1_AC*	58	1,306	TA	39	HGADGGGGCGTTCATTAAAG[a]	13.0 ± 0.62	7.73 ± 0.38
	TcMar-N1_AC	901	489	TA	203	CGAGGGCTATCCAGAAAGTA[a]	7.62 ± 0.55	4.39 ± 0.43
	TcMar-N2_AC	>1,000	783	TA	40	CGAGGGTTGAATGAAAAGTA[a]	6.93 ± 0.44	3.87 ± 0.27
	TcMar-N3_AC	>1,000	427	TA	24	CCSTGTTTCCCCGAAAATAA[a]	3.04 ± 0.41	1.45 ± 0.19
	TcMar-N4_AC	>1,000	661	TA	25	CYGTATATACTCGAGTATAA[a]	4.32 ± 0.27	2.32 ± 0.18
	TcMar-N5_AC	604	1,396	TA	23	CCSTGTTTCCCCGAAAATAA[a]	1.42 ± 0.12	0.79 ± 0.08
	TcMar-N6_AC	889	405	TA	23	CCGTATATACTCGAGTATAA[a]	11.1 ± 0.90	6.62 ± 0.57
	TcMar-N7_AC	>5,000	323	TA	24	CAGTAGAGTCTCACTTATCC[a]	2.91 ± 0.52	1.52 ± 0.38
	TcMar-N8_AC	>1,000	468	TA	25	CAGTAGAGTCTCACTTATCC[a]	8.39 ± 0.66	4.83 ± 0.41
	TcMar-N9_AC	105	339	TA	15	CAGTGYYCCCTCRCT	8.65 ± 0.91	5.54 ± 0.57
	TcMar-N10_AC	287	664	TA	311	CGAGGGCTATCCAGAAAGTT[a]	7.14 ± 0.55	5.18 ± 0.53
	TcMar-N11_AC	348	401	TA	15	CAGTGTTCCCTCRCT	10.1 ± 0.74	5.18 ± 0.53
	TcMar-N12_AC	952	321	TA	20	GAGTCTCRCTTATCCAACMT	10.5 ± 0.87	5.74 ± 0.57
	TcMar-N13_AC	36	168	TA	22	CYGTATTTCTTCAATTSTAA	22.6 ± 1.90	10.5 ± 0.88
	TcMar-N14_AC	14	240	TA	31	GATTGTAGCTACAGTATGAC[a]	27.3 ± 2.74	15.2 ± 1.67
	TcMar-N15_AC	498	524	TA	18	CASRKGTGTCAAACTCAA	23.1 ± 1.24	12.4 ± 0.76
	TcMar-N16_AC	3,771	350	TA	15	CAGTGTTCCCTCACT	3.02 ± 0.34	1.62 ± 0.19
	TcMar-N17_AC	137	356	TA	26	CAGGTTGAGYATCCCTTATC[a]	31.6 ± 1.61	17.1 ± 1.21
Chapaev	*Chapaev3-1_AC*	15	1910	TWA	12	CACTGRWAAACA	18.1 ± 0.99	12.3 ± 1.10
	Chapaev3-2_AC	4	1,767	TWA	18	CACTAGGAAACACAATTT	19.7 ± 1.46	10.9 ± 0.90
	Chapaev3-1N1_AC	46	390	TWA	11	CACWGSCCAAC	28.0 ± 2.39	15.9 ± 0.92
	Chapaev3-1N2_AC	988	320	TWA	19	CACTATGTAACAAAATTT	12.5 ± 0.77	6.00 ± 0.42

SE, standard error.
[a] Only the first 20 bp of the TIR are presented here.

Tc1/Mariner superfamily shows a more evenly distributed range of ages, from families that are extremely young and possibly active to much older families (up to 15% divergent from their consensus) that have long been extinct. Yet, young families seem to be predominant and, to a lesser extent than for *hATs*, the age distribution of *Tc1/Mariner* families is skewed toward young families. Similarly, *Helitron* families tend to be very young, in part because of recent lateral transfer, yet the presence of an ancient nonautonomous family suggests that they are not newcomers in the anole genome. The only exception to this pattern is the long-extinct *Chapaev* superfamily that contains only very divergent elements. The overall pattern of divergence of class 2 families indicate that DNA transposons readily reach fixation and accumulate in the anole genome as revealed by the continuous distribution of families diverging by 0–10%, surprisingly older (>10%) families are comparatively very rare and underrepresented in this genome. A possible explanation is that DNA transposons had a very low level of activity in the ancestral anole genome. However, it is surprising that the same pattern of low activity, followed by a more recent burst of activity, is shared by three unrelated superfamilies (*hAT*, *Mariner*, and *Helitron*). Another possibility is that transposons in anole are decaying rapidly by accumulation of indels and that they are no longer recognizable past a certain age. This rapid decay of mobile elements in the anole genome was previously reported for the *RTE Bov-B-1_AC* family of retrotransposons (Novick et al. 2009). This family's average divergence is only 4%, yet more than 70% of the full-length elements *RTE Bov-B-1_AC* have lost considerable amount of sequences through deletions. It is therefore not surprising that elements from families older than 10% are so fragmented that our approach failed to recognize them.

Class 2 transposons coexist in anole with a plethora of retrotransposon families. At least 42 families of non-Long Terminal Repeat (LTR) retrotransposons belonging to five clades are concurrently active in the anole genome (Novick et al. 2009). However, the dynamics of class 1 and class 2 elements differ drastically. The vast majority of non-LTR

retrotransposons families are much younger than class 2 families as they all have divergence lower than 2% (except 4 families of 46). In addition, retrotransposon families are on average far less numerous than class 2 families. It seems that, unlike class 2 elements, the vast majority of retrotransposon insertions does not reach fixation. This pattern is best explained by a high turnover of insertions resulting from a balance between the transposition of new elements and the selective loss of deleterious alleles. Interestingly, class 2 transposons do not seem to be subjected to the same level of purifying selection as class 1 elements. It was proposed that selection against non-LTR retrotransposons is length dependent and was caused, for the most part, by their ability to mediate ectopic recombination (Song and Boissinot 2007; Novick et al. 2009). Under this model, short elements are not eliminated by purifying selection to the same extent than long ones and, consequently, accumulate in the genome of their host. Indeed, most DNA transposons, in particular nonautonomous ones, are less than 2 kb long and much shorter than non-LTR retrotransposons. It is therefore likely that these elements are not as deleterious as retrotransposons and readily reach fixation. Interestingly, the longest class 2 transposons, the *Helitrons*, *hobo*, and *Polinton* elements (autonomous and nonautonomous), have a very low level of divergence and relatively low copy numbers, reminiscent of retrotransposons. It is therefore possible that those elements are subject to the same selective forces as non-LTR retrotransposons and are not reaching fixation, which could explain, in part, the lack of older *Helitron*, *Polinton*, and *hobo* related families.

At first glance, the diversity of DNA transposons in anole seems impressive, but how does it compares with other vertebrates? The analysis of teleostean genomes including salmonids, fugu, and zebrafish revealed that these genomes are also littered with DNA transposons, in particular from the *TC1/Mariner* superfamily (Krasnov et al. 2005). In fact, the diversity of fish DNA transposons far exceeds the one in lizards. For instance, the RepeatMasker output of the zebrafish, fugu, and stickleback genomes (available as tables at http://genome.ucsc.edu) reveals that teleostean genomes contain, in addition of the superfamilies we found in anole, several other superfamilies such as PiggyBac and MuDR. Similarly, the diversity and abundance of DNA transposons in *Xenopus tropicalis* is largely superior to the one in anole (Hellsten et al. 2010). In fact, class 2 transposons are the dominant category of repeated sequences in the frog genomes and account for about a third of its mass. In contrast, modern mammalian and avian genomes tend to lack active DNA transposons. In birds, DNA transposons have long been extinct, with the possible exception of a mariner family that was laterally transferred to the chicken genome (Wicker et al. 2005; Kordis 2009). Otherwise, the age and low abundance of class 2 transposons in birds are consistent with the paucity in repetitive sequences characteristic of bird genomes. Although class 2 elements appears extinct in most extent mammals (with the exception of recently recolonized genomes such as the bat and the mouse lemur; Ray et al. 2008; Pagan et al. 2010), class 2 elements have been diverse and prolific in early mammalian evolution. Mammalian genomes do contain a diversity of class 2 families even larger than the one reported in anole that accounts for as much as 3% of their size. However, these elements became independently extinct in multiple mammalian lineages around 40 million years ago (Lander et al. 2001; Waterston et al. 2002; Lindblad-Toh et al. 2005; Pace and Feschotte 2007). Thus, the diversity of active DNA transposons in the anole genome is reminiscent of the situation in ancestral mammalian genomes and could be used as a model to study the evolutionary processes that have shaped genomic evolution in the early stages of mammalian diversification.

Acknowledgments

We thank the Broad Institute Genome Sequencing Platform and Genome Sequencing and Analysis Program, Federica Di Palma, and Kerstin Lindblad-Toh for making the data for *A. carolinensis* available. We also thank Dr. Cedric Feschotte and two anonymous reviewers for their helpful comments on the manuscript. This research was supported by PSC-CUNY grant 61542-00-39 to Stéphane Boissinot.

Literature Cited

Arkhipova IR. 2006. Distribution and phylogeny of Penelope-like elements in eukaryotes. Syst Biol. 55:875–885.

Bao W, Jurka MG, Kapitonov VV, Jurka J. 2009. New super-families of eukaryotic DNA transposons and their internal divisions. Mol Biol Evol. 5:583–593.

Cordeaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 10:691–703.

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

Duvernell DD, Pryor SR, Adams SM. 2004. Teleost fish genomes contain a diverse array of L1 retrotransposon lineages that exhibit a low copy number and high rate of turnover. J Mol Evol. 59:298–308.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Edgar RC, Myers EW. 2005. PILER: identification and classification of genomic repeats. Bioinformatics 21:152–158.

Eickbush TH, Furano AV. 2002. Fruit flies and humans respond differently to retrotransposons. Curr Opin Genet Dev. 12:669–674.

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

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

Furano AV, Duvernell DD, Boissinot S. 2004. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. Trends Genet. 20:9–14.

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. 464:1347–1350.

Hartl DL, Lozovskaya ER, Lawrence JG. 1992. Nonautonomous transposable elements in prokaryotes and eukaryotes. Genetica. 86:47–53.
Hellsten U, et al. 2010. The genome of the Western clawed frog. Science. 328:633–636.
Hudson ME, Lisch DR, Quail PH. 2003. The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J. 34:453–471.
Jiao Y, Lau OS, Deng XW. 2007. Light-regulated transcriptional networks in higher plants. Nat Rev Genet. 8:217–230.
Jones JM, Gellert M. 2004. The taming of a transposon: V(D)J recombination and the immune system. Immunol Rev. 200:233–248.
Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A. 98:8714–8719.
Kapitonov VV, Jurka J. 2004. Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA Cell Biol. 25:311–324.
Kapitonov VV, Jurka J. 2005. Self-synthesizing DNA transposons in eukaryotes. Proc Natl Acad Sci U S A. 103:4540–4545.
Kordis D. 2009. Transposable elements in reptilian and avian (Sauropsida) genome. Cytogenet Genome Res. 127:94–111.
Kordis D, Gubensek F. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. Proc Natl Acad Sci U S A. 95:10704–10709.
Krasnov A, Koskinen H, Afanasyev S, Mölsä H. 2005. Transcribed Tc1-like transposons in salmonid fish. BMC Genomics. 12:107–117.
Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409:860–921.
Lerat E, Brunet F, Bazin C, Capy P. 1999. Is the evolution of transposable elements modular? Genetica. 107:15–25.
Le Rouzic A, Capy P. 2006. Population genetics models of competition between transposable element subfamilies. Genetics. 174:785–793.
Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H. 2007. Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. Science. 318:1302–1305.
Lindblad-Toh K, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 438:803–819.
Marchler-Bauer A, et al. 2007. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. 35:D237–D240.
Mathews S. 2006. Phytochrome-mediated development in land plants: red light sensing evolves to meet the challenges of changing light environments. Mol Ecol. 15:3483–3503.
Novick PA, Basta H, Floumanhaft M, McClure MA, Boissinot S. 2009. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. Mol Biol Evol. 26:1811–1822.
Novick PA, Smith J, Ray D, Boissinot S. 2010. Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. Gene. 449:85–94.
Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 17:422–432.
Pace JK, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A. 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.
Piskurek O, Austin CC, Okada N. 2006. Sauria SINEs: novel short interspersed retroposable elements that are widespread in reptile genomes. J Mol Evol. 62:630–644.
Piskurek O, Nishihara H, Okada N. 2009. The evolution of two partner LINE/SINE families and a full-length chromodomain-containing *Ty3/Gypsy* LTR element in the first reptilian genome of *Anolis carolinensis*. Gene. 441:111–118.
Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. Bioinformatics Suppl. 1:351–358.
Pritham EJ, Feschotte C, Wessler SR. 2005. Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. Mol Biol Evol. 22:1751–1763.
Ray DA, Feschotte C, Pagan HJ, Smith JD, Pritham EJ, Arensburger P, Atkinson PW, Craig NL. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. Genome Res. 18:717–728.
Robertson HM. 2002. Evolution of DNA transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): American Society Microbiology Press. p. 1093–1110.
Shedlock AM. 2006. Phylogenomic investigation of CR1 LINE diversity in reptiles. Syst Biol. 55:902–911.
Song M, Boissinot S. 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. Gene. 390:206–213.
Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circles transposons among animals. Genome Biol Evol. 2:656–664.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.
Waterston RH, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature. 420:520–562.
Wicker T, et al. 2005. The repetitive landscape of the chicken genome. Genome Res. 15:126–136.
Yang G, Nagel DH, Feschotte C, Hancock CN, Wessler SR. 2009. Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. Science. 325:1391–1394.

Associate editor: Bill Martin