# The Evolutionary Dynamics of Autonomous Non-LTR Retrotransposons in the Lizard *Anolis Carolinensis* Shows More Similarity to Fish Than Mammals

Peter A. Novick,\*† Holly Basta,‡ Mark Floumanhaft,\* Marcella A. McClure,‡ and Stéphane Boissinot\*†

\*Department of Biology, Queens College, The City University of New York; †Graduate School and University Center, The City University of New York; and ‡Department of Microbiology, Montana State University

The genome of the lizard *Anolis carolinensis* (the green anole) is the first nonavian reptilian genome sequenced. It offers a unique opportunity to comparatively examine the evolution of amniote genomes. We analyzed the abundance and diversity of non-LTR (long terminal repeat) retrotransposons in the anole using the Genome Parsing Suite. We found that the anole genome contains an extraordinary diversity of elements. We identified 46 families of elements representing five clades (L1, L2, CR1, RTE, and R4). Within most families, elements are very similar to each other suggesting that they have been inserted recently. The rarity of old elements suggests a high rate of turnover, the insertion of new elements being offset by the loss of element-containing loci. Consequently, non-LTR retrotransposons accumulate in the anole at a low rate and are found in low copy number. This pattern of diversity shows some striking similarity with the genome of teleostean fish but contrasts greatly with the low diversity and high copy number of mammalian L1 elements, suggesting a fundamental difference in the way mammals and nonmammalian vertebrates interact with their genomic parasites. The scarcity of divergent elements in anoles suggests that insertions have a deleterious effect and are eliminated by natural selection. We propose that the low abundance of non-LTR retrotransposons in the anole is related directly or indirectly to a higher rate of ectopic recombination in the anole relative to mammals.

#### Introduction

Non-LTR (long terminal repeat) retrotransposons, also known as retroposons (International Committee on Taxonomy of Viruses; Hull 2001), are autonomously replicating retroid agents that lack long LTRs. They have considerably affected the size, structure, and function of vertebrate genomes. This is exemplified by the fact that at least 30% of the genome of mammals is the result of their activity (Lander et al. 2001; Waterston et al. 2002). Although non-LTR retrotransposons were considered to be among the "junk DNA" class of repetitive elements, research has shown that they have been an extraordinary source of evolutionary novelties. Exaptation of these elements seems to have been relatively common during vertebrate evolution, either as part of coding sequences (i.e., exonization) or as regulatory elements (Makalowski 2000; Nekrutenko and Li 2001; Mikkelsen et al. 2007). In addition, the retrotransposition machinery encoded by non-LTR retrotransposons can act on other transcripts and is responsible for the amplification of SINEs (Short INterspersed Elements) and retroprocessed pseudogenes (Dewannieux et al. 2003; Dewannieux and Heidmann 2005). Some SINEs have also been co-opted as regulatory sequences and have played a major role in the early evolution of tetrapods (Bejerano et al. 2006).

Non-LTR retrotransposons constitute a diverse group of elements that are classified into 12 monophyletic clades (fig. 1) (Burke et al. 1999; Volff et al. 2000). These 12 clades diverged from each other more than 600 Ma and are found in most eukaryotes, including plants, animals, and fungi. Non-LTR retrotransposons encode for the proteins responsible for their own replication. They have a 5' untranslated region (UTR) that acts as an internal promoter

E-mail: stephane.boissinot@qc.cuny.edu.

*Mol. Biol. Evol.* 26(8):1811–1822. 2009 doi:10.1093/molbev/msp090 Advance Access publication May 6, 2009

© The Author 2009. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

and one or two open-reading frames (ORF), depending upon the clade. All non-LTR retrotransposons have a reverse transcriptase (RT) domain and 9 of the 12 clades encode for an apurinic/pyrimidinic endonuclease (Martin et al. 1995; Feng et al. 1996; Malik et al. 1999). The mechanism of insertion has not been resolved in all clades, but it is believed that non-LTR retrotransposons use a nick (generated by their own endonuclease) on the host chromosome to prime reverse transcription of the RNA transcript directly into the target site (Luan et al. 1993; Luan and Eickbush 1995; Cost et al. 2002). The majority of new elements produced by this mechanism are truncated at their 5' end and incapable of further retrotransposition ("dead-on-arrival"). Elements belonging to different clades can be further subdivided into families based on shared differences between elements (Furano 2000). As the vast majority of elements do not serve a function for the host, they accumulate mutations at the neutral rate, so that old families of elements are more divergent than younger ones (Voliva et al. 1983; Pascale et al. 1993).

The analysis of complete genome sequences has revealed some striking differences in the abundance and diversity of non-LTR retrotransposons (Lander et al. 2001; Waterston et al. 2002; Volff et al. 2003; Duvernell et al. 2004; Furano et al. 2004; Gentles et al. 2007; Warren et al. 2008). In placental mammals, a single clade, LINE-1 (Long Interspersed Nuclear Element, L1), has remained active since the split between eutherians and marsupials. L1 has been very successful in eutherians, reaching extremely high copy numbers (e.g., 500,000 copies in the human genome) (Lander et al. 2001; McClure et al. 2005). Most of these elements are ancient and are the product of past amplifications. Phylogenetic analyses have shown that most modern mammalian genomes contain a single dominant lineage of L1 families, suggesting that the evolution of L1 in mammals is controlled in ways that prevent further diversification (Furano 2000; Furano et al. 2004). The reason why L1 evolves as a single lineage is unknown, but competition between elements for host-encoded transcription factors could explain the unusual evolution of L1 families (Khan et al. 2006).

Key words: Anolis carolinensis, long interspersed nuclear element, retroposon, retrotransposon.



FIG. 1.—Phylogenetic relationships among the 12 clades of non-LTR retrotransposons. *Anolis* consensus sequences are framed in gray. This NJ tree was constructed from a portion of the translated RT domain. Bootstrap values less than 75 have been removed. Numbers next to non-Anolis sequences correspond to Genbank accession numbers listed in the Materials and Methods section. The letters "AC" stand for *Anolis carolinensis*.

In contrast, the genome of teleostean fish contains several active clades, each represented by divergent families (Volff et al. 2000; Duvernell et al. 2004; Furano et al. 2004; Neafsey et al. 2004; Basta et al. 2007). For instance, 32 divergent L1 families coexist in zebrafish, each represented by small numbers (<50) of very similar elements (Duvernell et al. 2004; Furano et al. 2004). The differences in diversity and copy number between fish and mammals could result from a number of factors, including differences in the control of retrotransposition by the host, competitive interactions between families of elements, variations in the intensity of selection against new inserts, and the history of populations. The respective importance of these different factors is unclear and remains a matter of debate (Eickbush and Furano 2002; Furano et al. 2004; Neafsey et al. 2004; Kordis et al. 2006; Song and Boissinot 2007).

The reduction in retrotransposon diversity and the explosion in copy number of the L1 clade in mammals represent one of the major transitions in the evolution of vertebrate genomes. Until recently, it was difficult to study this important question because of a lack of genomic data in amphibians and reptiles. The recent completion of the lizard (Anolis carolinensis) and frog (Xenopus tropicalis) genomes bridges a gap between fish and mammals and will provide new insights into the evolution of tetrapod genomes. Anolis carolinensis, the green anole, is a small lizard (Squamata: Iguanidae) found in the southeast United States that has become an important model organism in evolutionary and behavioral studies. We found that the anole genome contains at least five active clades of non-LTR retrotransposons that vary considerably in replicative success and diversity. Within most families, elements are very similar to each other suggesting that they have been recently inserted. The lack of elements of intermediate or old age relative to young elements indicates that retrotransposons accumulate in the anole genome at a low rate, possibly because the deleterious impact of retrotransposons is stronger in anoles than it is in mammals.

# Materials and Methods

## Sequence Acquisition

The results reported here are from the Genome Parsing Suite (GPS) software (McClure et al. 2005) used for identification and classification of A. carolinensis retroid content. Of all the retroid components, the RT is the best conserved through evolutionary time (McClure et al. 1988) and is essential for autonomous transposition. The GPS analysis is therefore centered on the RT, identifying it first and then expanding to other components. The approach of the GPS is radically different from Repeat Masker, which is used to mask out and count repetitive agents using consensus DNA sequences (Smit et al. 2004). Repeat Masker and similar methods suffer from the loss of signal due to mutational saturation because DNA is used to query a genome rather than amino acid sequences. DNA sequence libraries are also often unable to detect new components.

The *A. carolinensis* genome v. 1.0 was downloaded from the University of Santa Cruz Genome Bioinformatics Web site (Karolchik et al. 2003). The *A. carolinensis* genome has 1.7 Gb of its expected 2.2 Gb sequenced (about 77%); therefore, any numbers presented in this study are pending further genome refinement. The GPS was populated with 130 Retroid agent queries in this scan of the A. carolinensis genome. Although only the non-LTR retrotransposons found in full length (FL) in the A. carolinensis genome are analyzed in depth in this study, queries representative of other Retroid families were included to ensure correct classification of all Retroid agents. The queries include sequences that are specific to humans, birds, Anolis, Xenopus, and fish, in addition to a set of 30 queries that represent the major families of all Retroid agents. As little has been classified in the Anolis genome, novel Anolis queries were selected from the GPS results using the following criteria: completeness (presence of the most components), containing the least number of stop codons and frameshifts, containing all the motifs of the enzymatic core proteins, and phylogenetically representing their family. Phylogenetic trees were constructed using all the FL copies of each family to ensure that outliers were not selected as queries.

The GPS method is divided into two stages: stage I GPS deals solely with the RT, whereas stage II GPS classifies FL agents. In Stage I GPS, Washington University Basic Local Alignment Search Tool translated nucleotides (WU-TBlastN) version 2.0 (Gish 1996–2004) was used to query the A. carolinensis genome with the following parameters: E = 1, -matrix pam70, Q = 9, R = 1, V =1e7, B = 1e7, gapL = 0.307, gapK = 0.13, gapH = 0.7,X = 15, gapX = 33, gapW = 44, gapS2 = 63, S2 =41, hspmax = 0, and -span. After WU-TBlastN scans, the A. carolinensis genome with the Retroid RT queries, Stage I GPS sorts and filters raw WU-TBlastN hits, which are redundant and contain false positives, due to: 1) alternative alignments for a given query to a specific region, 2) cross coverage of the queries, and 3) counting as unique, a number of small hits that are actually from the same gene. After sorting by query, chromosome, polarity, and reading frame, the GPS compounds small hits and removes false positives due to cross coverage on these compounded hits. The GPS removes redundancy by deleting hits that are completely covered by a longer hit to the same position, thereby preventing overestimation of the amount of potential RT genes. Single contiguous sequences, single compound hits composed of subsequences, and sets of ambiguous hits to the same position and reading direction are all considered unique RT hits. Ambiguous cases are often resolved in Stage II of the GPS. Unique hits are then assessed for quality first by degree of Ordered Series of Motifs conservation (McClure 1991), which is made up of six highly conserved motifs that fold to form the active site of the enzyme (Kohlstaedt et al. 1992), and then by presence of frameshifts and stop codons. FL RT hits with neither frameshifts nor stop codons are labeled "perfect." In Stage II, GPS, each RT found by Stage I, has its position in the host's genome extended 7 kb upstream and downstream, creating a 14 kb (plus the size of the RT) cutout. Using this RT-outward approach, the GPS is able to construct potential Retroid agent genomes. WU-TBlastN is used a second time to compare each 14kb+ cutout with the RT hit's corresponding component library. If an RT hit is ambiguous between multiple queries, then each of these queries' component libraries are searched, and the query with the highest score over all components is called as the closest to the new Retroid agent. If the sequence has "all" the gene components in the query's genomic order it is considered FL. All genomes with one frameshift or stop codon, as well as those that are error free (perfect), are considered to be potentially active, because Retroid agents are known to overcome such mutational errors by translational recoding (http://recode.genetics.utah.edu/) thereby producing functional proteins. Note that some queries themselves may contain frameshifts and stop codons. For a more in-depth discussion of the GPS, see McClure et al. (2005) and Basta et al. (2007). When GPS finds a sequence that has identity to known non-LTR retrotransposon components, but no UTRs, and these regions do not pull out any known UTRs when Blast searches are executed (Altschul et al. 1990), a test for novel UTRs is performed. When multiple copies of the sequence are available, 1-kb 5' is aligned to other 5' regions from additional copies, and 3' regions are aligned to other 3' regions, and conserved regions are considered UTRs.

#### Characterization of Retrotransposon Families

Extracted sequences were aligned using ClustalW in the program BioEdit (Hall 1999) and subsequently categorized into families and subfamilies based upon sequence similarity. Branching patterns of phylogenies created by Neighbor Joining (NJ) under the Kimura 2-paramaters' distances in the program MEGA 4.0 (Tamura et al. 2007) was used in tandem to make sure a family had not been overlooked. A sequence from each family was then submitted to Repeat masker (www.repeatmasker.org) to verify proper clade classification. In order to determine the ancestral progenitor, FL consensus sequences were constructed for each family. Consensus sequences have been deposited in Repbase (http://www.girinst.org/repbase). Each consensus sequence was then used in a BLAT search and the resultant output sequences were collected for further analysis. FL copy numbers were determined from the original GPS output. Copy number of truncated elements that contained less than 200 copies was retrieved from the aforementioned BLAT output, whereas estimates for larger families were calculated using the electronic polymerase chain reaction option as well as from the original GPS output. FL consensus sequences from each family were compared with each other using dotmatcher (Rice et al. 2000). Additionally, each consensus sequence was analyzed for their individual GC content using the sequence analysis option in DAMBE (Xia and Xie 2001). The genomic environment for at least 20 sequences in each family was also characterized. Of 5' and 3' flanks, 50 kB was collected from the genome browser and analyzed for their GC content using DAMBE. In order to determine the relative time of amplification of each family, at least 500 bp of the RT domain of at least 20 elements were aligned and analyzed by MEGA 4.0, and both pairwise divergence and average from the consensus were calculated using Kimura's 2-parameter's distance.

#### **Evolutionary Relationships**

Consensus sequences for each family were submitted to NCBI's ORF Finder and Conserved Domains (Marchler-Bauer et al. 2007) to find the location and frame in which the ORFs were located. ORFs were extracted and aligned in nucleic acid and amino acid and the size of each was recorded. Previously published RT domains of each clade were recovered and aligned using ClustalW along with the aforementioned novel proteins. Sequence numbers correspond to those in figure 1 and can be found under the accession numbers: 1—AF086712; 2-L25662; 3-X99080; 4—AF018033; 5—M93690; 6—X51968; 7-D38414; 8-AF012049; 9-X06950; 10-M22874; 11—M14954; 12—M28878; 13—X60372; 14—U66331; 15—U88211; 16—AB005891; 17—U73800; 18-X60177: 19—U87543; 20—Z25525; 21—AF025462; 22—U58755; 23—M26915; 24—AF081114; 25—U93574; 26—X51967; 27—U13035; 28—M16558; 29—U29445; 30-M33009; 31-M62862; and 32-X17078. These sequences were then analyzed phylogenetically using the NJ and maximum likelihood tree using MEGA 4.0 and PHYML online (Guindon et al. 2005), respectively.

#### Results

The GPS identified a total of 1,888 FL elements in the anole. Further Blast and BLAT searches of the anole genome using the sequences collected by GPS as probes failed to detect any other FL elements, although we recovered a number of truncated (TR) elements (supplementary fig. 1, Supplementary Material online). Representatives from five of the twelve clades of non-LTR retrotransposons were detected: L1, L2, CR1, RTE, and R4 (fig. 1). These clades differ greatly in copy number and diversity.

Of the five clades that inhabit the genome of A. carolinensis, L1 is the least numerous, with 170 FL and 626 TR elements, yet it is the most diverse. A phylogenetic analysis based on the two ORFs (fig. 2) reveals the presence of 20 distinct L1 families, represented by very low copy number (7-144 elements; table 1). A group of elements (L1-like in fig. 1) structurally similar to L1 did not branch with other L1 families, but instead clustered with the RTE clade (with very low bootstrap support). As the structure of these elements suggests an evolutionary affinity to L1, they were provisionally classified as L1like. L1 families are very divergent from each other and their separation predates the split between reptiles and mammals (fig. 1). Within each family, FL elements appear very closely related to each other (fig. 2). The phylogenetic tree also shows a near complete absence of internal branches indicating a lack of old L1 inserts in the anole. In addition, repeated Blast and BLAT searches of the anole genome using the 3' UTR as a probe failed to detect any divergent TR L1 elements. The recent origin of most L1 inserts is confirmed by the very low level of divergence within families (<1% divergence from their consensus; fig. 3A and table 1) for both TR and FL elements. The fact that 50% of FL elements have both ORFs intact (table 2) is also consistent with the young age of these elements. Therefore, it seems that the vast majority of L1 elements inserted very recently in the anole genome. The near absence of old L1 inserts indicates that the vast majority of L1 elements do not reach fixation in the anole genome,



FIG. 2.—Phylogenetic relationships among the 20 L1 anole families recovered from the GPS output. The tree was constructed using NJ and bootstrap values >75% are shown. Arrows indicate the acquisition of novel 5' UTRs.

although L1 elements are still active and are probably ancient residents of this genome.

The L2 clade is also very diverse and is represented by 17 monophyletic families (fig. 4), ranging in copy number from 71 to 529. L2 families are divergent from each other although not as much as L1 families, the deepest node in the L2 clade corresponding to approximately one-fifth of the deepest divergence in the L1 clade (fig. 1). The majority

Table 1	
Characteristics of Non-LTR Retrotransposons Present in the Anole Genome Separated by Fan	nily

Clade	Family	FL Size (bp)	FL Copy Number	Truncated Copy Number	GC Element	GC Environment	Nb. AAs ORF1	Nb. AAs ORF 2	Average Pairwise Divergence $(\% \pm SE)^a$	Average Divergence from Consensus $(\% \pm SE)^a$
CR1	CR1 AC 1	5,300	47	593	46.43	40.55	423	911	$0.38 \pm 0.07$	$0.13 \pm 0.01$
	CR1 AC 2	5,843	25	316	46.66	40.21	423	911	$0.28 \pm 0.05$	$0.07 \pm 0.03$
	CR1 AC 3	5,708	37	467	46.61	40.38	423	911	$1.12 \pm 0.12$	$0.51 \pm 0.12$
	CR1 AC 4	4,653	8	101	46.52	40.33	423	911	$2.89 \pm 0.34$	$1.14 \pm 0.15$
L1	L1 AC 1	6,378	5	5	37.19	40.11	353	1,269	$0.19 \pm 0.20$	$0.10 \pm 0.04$
	L1 AC 2	6,305	7	4	39.02	40.11	353	1,335	$0.57 \pm 0.23$	$0.46 \pm 0.14$
	L1 AC 3	6,300	11	10	38.14	40.11	353	1,335	$0.62 \pm 0.22$	$0.39 \pm 0.14$
	L1 AC 4	6,206	3	4	37.58	40.11	353	1,267	$0.33 \pm 0.09$	$0.12 \pm 0.07$
	L1 AC 5	6,190	3	5	38.08	40.11	353	1,259	$0.88 \pm 0.33$	$0.00 \pm 0.00$
	L1 AC 6	6,153	7	10	36.47	39.28	353	1,250	$0.35 \pm 0.11$	$0.17 \pm 0.07$
	L1 AC 7	6,433	4	8	35.69	40.16	354	1,250	$0.21 \pm 0.10$	$0.13 \pm 0.40$
	L1 AC 8	6,430	4	15	35.41	40.16	354	1,265	$0.42 \pm 0.15$	$0.18 \pm 0.08$
	L1 AC 9	6,410	7	98	35.80	39.20	341	1,251	$1.10 \pm 0.30$	$0.50 \pm 0.11$
	L1 AC 10	6,848	4	8	36.05	40.58	359	1,246	$0.33 \pm 0.12$	$0.25 \pm 0.09$
	L1 AC 11	6,648	6	11	36.54	40.58	366	1,256	$0.00 \pm 0.00$	$0.00 \pm 0.00$
	L1 AC 12	6,500	5	9	34.80	38.89	353	1,270	$0.55 \pm 0.07$	$0.27 \pm 0.03$
	L1 AC 13	6,758	3	8	35.29	40.58	361	1,251	$0.35 \pm 0.11$	$0.17 \pm 0.05$
	L1 AC 14	6,712	4	8	35.71	40.58	361	1,251	$0.53 \pm 0.14$	$0.41 \pm 0.11$
	L1 AC 15	6,162	11	75	37.51	39.52	349	1,243	$0.87 \pm 0.21$	$0.42 \pm 0.09$
	L1 AC 16	6,130	17	75	34.31	38.69	349	1,260	$1.20 \pm 0.16$	$0.54 \pm 0.08$
	L1 AC 17	5,410	18	85	35.86	38.91	351	1,252	$0.70 \pm 0.07$	$0.35 \pm 0.03$
	L1 AC 18	5,524	24	120	35.60	39.17	374	1,240	$1.36 \pm 0.26$	$0.63 \pm 0.04$
	L1 AC 19	5,758	4	16	36.49	37.87	351	1,245	$1.63 \pm 0.26$	$0.76 \pm 0.13$
	L1 AC 20	5,250	23	52	33.53	38.98	380	1,243	$0.66 \pm 0.14$	$0.35 \pm 0.06$
	L1-Like AC 1	5,211	10	200	41.34	40.29	483	1,035	$1.79 \pm 0.93$	$0.85 \pm 0.48$
L2	L2 AC 1	6,058	15	230	50.17	38.88	320	1,001	$0.41 \pm 0.07$	$0.29 \pm 0.02$
	L2 AC 2	5,890	30	172	50.08	38.68	318	1,001	$1.40 \pm 0.23$	$0.98 \pm 0.15$
	L2 AC 3	4,980	15	155	50.41	38.46	318	785	$1.63 \pm 0.21$	$1.27 \pm 0.08$
	L2 AC 4	6,077	15	80	53.51	38.40	341	785	$0.50 \pm 0.10$	$0.40 \pm 0.08$
	L2 AC 5	5,410	30	146	52.55	38.27	335	786	$1.05 \pm 0.11$	$0.98 \pm 0.07$
	L2 AC 6	6,012	24	180	53.51	39.24	278	785	$0.86 \pm 0.11$	$0.48 \pm 0.04$
	L2 AC 7	5,200	30	300	52.11	38.58	317	785	$0.43 \pm 0.10$	$0.31 \pm 0.05$
	L2 AC 8	5,890	30	200	53.76	38.77	317	889	$0.61 \pm 0.11$	$0.47 \pm 0.03$
	L2 AC 9	5,765	22	50	51.74	39.44	316	855	$0.58 \pm 0.15$	$0.31 \pm 0.03$
	L2 AC 10	5,700	21	508	52.89	39.11	342	785	$1.47 \pm 0.16$	$1.11 \pm 0.07$
	L2 AC 11	6,288	18	315	51.57	39.17	342	785	$0.62 \pm 0.09$	$0.40 \pm 0.03$
	L2 AC 12	4,823	15	220	51.03	38.31	341	792	$1.47 \pm 0.39$	$1.33 \pm 0.12$
	L2 AC 13	5,052	49	170	50.59	37.70	343	785	$1.03 \pm 0.17$	$1.00 \pm 0.06$
	L2 AC 14	5,110	18	144	51.66	39.18	333	798	$1.71 \pm 0.16$	$1.08 \pm 0.10$
	L2 AC 15	5,960	18	400	52.04	38.59	332	919	$0.73 \pm 0.07$	$0.43 \pm 0.04$
	L2 AC 16	5,970	18	75	49.76	38.92	NA	696	$4.74 \pm 0.36$	$2.86 \pm 0.17$
	L2 AC 17	5,060	12	75	49.88	38.33	362	988	$2.59\pm0.26$	$2.02 \pm 0.14$
R4	R4 AC 1	3,760	365	734	42.57	40.14	NA	1,137	$1.33 \pm 0.10$	$0.41 \pm 0.03$
	R4 AC 2	3,760	264	532	42.57	40.14	NA	1,138	$1.51 \pm 0.12$	$0.39 \pm 0.05$
RTE	RTE Bov-B AC 1	3,226	61	3203	45.57	39.89	NA	931	$3.90\pm0.36$	$0.24 \pm 0.09$
	RTE-1 AC 1	3,910	156	96	50.58	40.49	NA	1,081	$0.18 \pm 0.08$	$0.07 \pm 0.02$

<sup>a</sup> Divergence was calculated using the Kimura 2-parameter correction in MEGA 4.0.

of L2 families show evidence of recent activity and are represented by young elements. For instance, the L2AC1 family contains only elements that diverge from each other by 0.41% on average, suggesting that this family might still be active (fig. 3*B*). We did not detect any L2AC1 element diverging from the family consensus by more than 1%, suggesting that, like L1 elements, L2AC1 elements are not accumulating (i.e., reaching fixation) in the anole genome. In fact, of 17 L2 families, 15 families have an average divergence from consensus lower than 1%. This is consistent with the fact that 44.6% of all FL L2 elements have an intact ORF2 (table 2). Only two families have an average divergence higher than 2% (L2AC16 and L2AC17). Family L2AC16 has an average divergence of 4.74% and the lowest divergence between two L2AC16 elements is 2.76% (table 1). This indicates that this family has not been recently active and contains elements that are likely to be fixed. Interestingly, the two L2 families that show evidence of accumulation are also the least numerous ones in our collection. The dynamics of amplification of these L2 families is reflected in the distribution of pairwise divergence between elements (fig. 3*B*): At some point in time, these families began accumulating in small numbers in the anole genome and their accumulation coincided or was soon followed by their decline and likely extinction. It is interesting to note that the only two L2 families that show evidence of accumulation are also the only ones that seem to be no longer active. However, these two families are exceptions and the vast majority of L2 elements are very young.



FIG. 3.—Pairwise divergence distribution of families belonging to all five clades recovered from the anole genome: A—L1 families; B—L2 families, C—CR1 and R4 families; and D—RTE families.

The CR1 and R4 clades show very little family structure compared with L1 and L2, yet they are found in high copy number, reaching  $\sim$ 1,600 and  $\sim$ 3,000 copies, respectively. Only four CR1 families are supported by the phylogenetic analyses (fig. 5A). The R4 clade shows a little more diversity than CR1 but more than half of all R4 elements belong to two major families (R4AC1 and R4AC2 in fig. 5B). CR1 and R4 families are not as divergent as L1 and L2 families. For instance, the deepest node on the CR1 tree corresponds to 4.3% divergence. Most CR1 and R4 inserts are recent as suggested by the low divergence between elements within each family (<1% from consensus) and by the high proportion of FL elements with intact ORF (table 2). The CR1 copy number reported here contrasts with previous estimates of ~300,000 CR1derived sequences in the anole genome (Shedlock et al. 2007). Indeed, a screening of the anole genome using GPS revealed a large number of small DNA segments derived principally from CR1 sequences ("total RT hits" in table 2). However, most of these elements are highly fractioned and degenerate suggesting they are very ancient.

The RTE clade contains two very divergent lineages (RTE Bov-B AC and RTE-1 AC) that separated before the

origin of vertebrates (Zupunski et al. 2001). The Bov-B family is one of the most successful non-LTR retrotransposons in the anole, with 3,325 copies, whereas the RTE-1 family is just under 250 copies. The phylogeny of Bov-B and RTE-1 elements did not reveal any clear subsets (data not shown) like we observed for other clades, suggesting that a small number of closely related progenitors are responsible for the amplification of RTE elements. RTE-1 elements are extremely similar to each other, with an average divergence of 0.21%, suggestive of their young age. In contrast, the average divergence of the Bov-B family is 3.9%. The pairwise divergence distribution of the Bov-B family (fig. 3D) shows that this family has not produced recent insertions and is probably no longer active, although 21% of FL Bov-B elements have an intact ORF. The large divergence of this family suggests that the vast majority of Bov-B inserts is likely to be fixed. The pairwise divergence distribution of Bov-B (fig. 3D) is similar to the one of L2AC16 and L2AC17 and provides another example of an extinct family that accumulated in the anole genome.

Because the Bov-B family is relatively ancient and numerous, it offers the opportunity to examine the decay of elements in the anole. Using the first 150 bp of a Bov-B

 Table 2

 Characteristics of Non-LTR Clades in the Anole Genome

Clade	No. FL	No. Total	No. Total RT Hits	% ORF 1 <sup>a</sup>	% ORF 2 <sup>a</sup>	% ORF 1 and $2^{a}$
L1 AC	180	1,006	7,441	65.39%	80.77%	50.00%
L2 AC	380	3,800	38,607	20.05%	44.61%	9.52%
CR1 AC	117	1,594	86,802	26.50%	39.32%	14.53%
R4 AC	994	3,000	7,682	Х	31.99%	Х
RTE Bov-B AC <sup>b</sup>	61	3,264	17,224	Х	21.34%	Х
RTE-1 AC <sup>b</sup>	156	2,52	1,330	Х	35.35%	Х

<sup>a</sup> Percent of FL elements with ORFs intact was calculated from the original GPS output.

<sup>b</sup> RTE Bov-B AC and RTE-1 AC were treated as separate clades due to their divergence and differences in evolutionary dynamics.



Fig. 4.—Phylogenetic relationships among the 17 L2 anole families recovered from the GPS output. The tree was constructed using NJ and only bootstrap values >75% are shown. Arrows indicate the acquisition of novel 5' UTRs.

element as a probe, we performed a BLAT search of the genome. Each hit was retrieved together with 7 kb of downstream DNA and aligned to the anole Bov-B consensus sequence. As elements with an intact 5' end are presumed to be FL, we should be able to find a 3' UTR about 3 kb downstream of each 5' terminus. In fact, we did not find a 3' end downstream of most 5' sequences, as only 28% of elements extended all the way from the 5' UTR to the 3' end (fig. 6). The same analysis performed on the much younger RTE-1 family revealed that 65% of 5' terminal sequences extend all the way to the 3' end. For comparison, we performed the same analysis in human using the 5' end of a L1PA6 and of a L1PA11 element. The average divergence of the LPA6 family is the same as Bov-B AC ( $\sim 4\%$ ), whereas L1PA11 family is much older and is about 16% divergent (Khan et al. 2006). Here we found the 3' end corresponding to the 5' end for 86% and 78% of L1PA6 and L1PA11 elements, respectively, although L1 elements are about twice as long as RTE elements. This indicates that elements decay much faster in anoles than in humans. FL elements that do not extend to their 3' extremity not only contain small deletions, but they also completely miss their 3' end. This suggests that the decay of FL elements is probably due to large deletions, possibly mediated by ectopic recombination between elements, instead of DNA loss by small deletions. The rapid decay of elements explains, in part, why we identified very few old (i.e., elements that diverge by more than 5% from their consensus) FL elements.

All the elements we collected are typical members of their clade. However, comparisons of consensus sequences revealed a striking difference between FL elements in the L1 and L2 clades. Although the coding region of these families is quite conserved within each clade and among families within clades, the 5' UTR is not. In fact, we identified 15 and 16 different 5' UTRs in the L1 and L2 clades, respectively, that account for differences in the length of consensus elements in table 1. Although the first 20–40 bps are shared across L1 families, these 5' UTR sequences show very little homology to each other. For instance, figure 7 shows a dot plot comparison of two L1 and L2 consensus sequences. These consensus sequences differ by less than 10% outside the 5' UTR and can be easily aligned, yet these sequences have completely different, that is nonhomologous, 5' UTRs. Interestingly, the family diversity of each clade correlates nicely with the diversity in 5' UTR sequences, each major L1 and L2 lineages have a different 5' UTR (figs. 2 and 4). In contrast, clades with low family diversity, like RTE, CR1, and R4, do not show any diversity at their 5' end. This correlation suggests that the ability to recruit novel promoter sequences in L1 and L2 drives the evolution of simultaneously active families and might be responsible for the diversity of these clades.

#### Discussion

The anole genome is the first nonavian reptilian genome whose complete sequence is available. It bridges a large phylogenetic gap and provides a unique opportunity to comparatively investigate the evolution of amniote genomes. We found that the anole genome contains an extraordinary diversity of non-LTR retrotransposons. Five clades show signs of very recent activity and two of these clades, L1 and L2, contain numerous families, some of which have been simultaneously active since the separation between mammals and reptiles (315 Ma). We identified at least 46 recently active families of non-LTR retrotransposons in the anole and we estimate the number of potentially active elements (i.e., FL elements with both ORF intact) to be about  $\sim$ 500 copies. This situation is reminiscent of the diversity of non-LTR retrotransposons in teleostean fish. Fish genomes usually contain several active clades of non-LTR retrotransposons, sometimes represented by a large diversity of families (Volff et al. 2003; Duvernell



Fig. 5.—Phylogenetic relationships among R4 (A) and CR1 (B) elements recovered from the GPS output. The tree was constructed using NJ and bootstrap values >75% are shown. The arrow indicates the acquisition of a novel 5' UTRs.

et al. 2004; Furano et al. 2004; Neafsey et al. 2004; Basta et al. 2007). For instance, the pufferfish (*Tetraodon nigro-viridis*) and the zebrafish (*Danio rerio*) genomes contain six and seven active clades of non-LTR retrotransposons, respectively (Volff et al. 2003; Basta et al. 2007). In zebrafish, the L1 clade is represented by 32 active families that diverged before the origin of vertebrates (Furano et al. 2004). This situation contrasts greatly with mammalian ge-

nomes that are dominated by a single clade, L1 in eutherians and marsupials (Furano et al. 2004; Gentles et al. 2007), and L2 in monotremes (Gilbert and Labuda 2000; Warren et al. 2008). L1 usually evolves in mammals as a single lineage, so that only a unique family of closely related elements is active at a given time (Furano 2000; Furano et al. 2004). For instance, a single family, called Ta, is currently active in human (Skowronski et al.



FIG. 6.—Length of elements as they extend from the 5' end. At least 80 sequences from each of the 4 families, RTE Bov-B AC (*A*), RTE-1 AC (*B*), L1PA11 (*C*), and L1PA6 (*D*).

1988). This family contains only two subsets of active elements, Ta-0 and Ta-1, which are considerably less divergent (<1%) from each other than anole's L1 families are (Boissinot et al. 2000).

The much greater diversity of non-LTR retrotransposons in the anole and in teleostean fish does not translate into larger genome sizes. In fact, mammalian genomes are significantly larger than those genomes. The human and mouse genomes are 3.2 and 2.8 Gb, respectively; in comparison, the *Anolis* genome is only 2.2 Gb, whereas teleostean genomes vary between 0.4 Gb (in *T. nigroviridis*) and 1.7 Gb (in *D. rerio*) (Volff et al. 2003). The larger size



FIG. 7.—Dotmatcher results comparing the 5' UTR of two closely related L1 (Left) and L2 (Right) consensus sequences.

of mammalian genomes relative to other vertebrates is directly related to the abundance of L1 elements in mammals. It was estimated that the human genome contains  $\sim$  500,000 copies of L1 elements that account for  $\sim 17\%$  of our DNA (Lander et al. 2001). In contrast, we identified only ~16,000 non-LTR retrotransposons (FL and TR), accounting for  $\sim 1.3\%$  of the anole genome. In addition to the recent elements we characterized, the anole genome contains a large number of small DNA fragments, between  $\sim$ 160,000 (based on the number of RT hits; table 2) and 300,000 (Shedlock et al. 2007) that are principally the product of ancient CR1 activity. These ancient elements are more numerous in anole than in chicken ( $\sim 200,000$ ; Shedlock et al. 2007) and in zebrafish ( $\sim$ 102,000; Basta et al. 2007). When these old degenerate elements are taken into account, the total fraction of the anole genome derived from non-LTR retrotransposons is still largely inferior  $(\sim 10\%)$  to that of mammals. Therefore, the accumulation of non-LTR retrotransposons found in mammals is truly specific of this class of vertebrates and suggests a major difference between mammalian and nonmammalian vertebrates in the way host genomes interact with their parasitic non-LTR retrotransposons.

The scarcity of divergent elements and the abundance of very young inserts indicate that the vast majority of non-LTR retrotransposons do not reach fixation in the anole genome. This suggests a rapid turnover of elements, in which the insertion of new elements is offset by the loss of element-containing loci. This mode of evolution is similar to the turnover of elements in Drosophila, where selection against element-containing loci limits copy number (Charlesworth and Langley 1989; Eickbush and Furano 2002). Similarly, the turnover of elements in the anole suggests that selection against retrotransposon inserts must be strong enough to prevent their accumulation. L1 inserts, in particular long ones, are also negatively selected in mammals (Boissinot et al. 2001, 2006; Song and Boissinot 2007), yet they do accumulate in mammalian genomes. Therefore, the cost that non-LTR elements impose on the fitness of their host is likely to be higher in the anole than it is in mammals. The deleterious effect of retrotransposons can result from their ability to mediate ectopic recombination (Langley et al. 1988), effects on gene activity (Charlesworth and Charlesworth 1983), or the retrotransposition mechanism per se (Boissinot et al. 2001). Because the five clades are all characterized by a high rate of turnover, non-LTR retrotransposons must have a common deleterious effect that is independent of their clade and that affect equally FL and truncated elements.

A deleterious effect due to the retrotransposition process is unlikely because selection should prevent the fixation of retrotransposition-competent FL elements but not of TR elements because they are incapable of further retrotransposition. Gene inactivation is also an unlikely mechanism because the anole genome is sufficiently large and contains regions with low gene density where non-LTR retrotransposons could accumulate. Chromosomal rearrangements and large DNA deletions caused by ectopic (i.e., nonallelic) recombination predicts that selection should act primarily against long elements because they are more likely to be involved in ectopic exchange than shorter ones. Experimental and genomic evidence in mammals suggests that ectopic recombination occurs rarely if the length of sequence homology is shorter than 1.2 kb (Cooper et al. 1998; Song and Boissinot 2007). As short elements are also rare in the anole genome, it is doubtful that this mechanism plays a major role, unless ectopic recombination in reptiles requires a much shorter length of homology than in mammals or unless the rate of recombination is much higher in anole than in mammals (see below). However, this model fails to explain why many elements belonging to some specific families (L2AC16, L2AC17, and Bov-B) have reached fixation in the anole genome, as there is no reason to believe that these elements were less likely to mediate ectopic recombination. Recent experimental work suggests a possible extension of the ectopic exchange model. In yeast, a defense mechanism limits interelement recombination by changing the conformation of the chromatin at the insertion site and in neighboring sequences (Ben-Aroya et al. 2004). If chromatin modification has a negative impact on the function of the genome, all elements, whatever their length or their clade, would be deleterious and therefore eliminated by selection. If some families lack sequence motifs recognized by the surveillance machinery of the host, they could temporarily evade host defense and accumulate. This model is still speculative because a defense mechanism, designed to prevent interelement recombination, has yet to be found in vertebrates.

Our analysis of the Boy-B family emphasizes another mechanism that could account for both the smaller size of the anole genome and for the scarcity of old elements. We found that elements in the anole decay much faster than their mammalian counterpart and that their decay results from the loss of their ends presumably caused by interelement recombination. This pattern is very similar to the decay of CR1 elements in chicken (Abrusan et al. 2008) where the low abundance of elements could be the consequence of their high recombination rate. Although very little is known about the recombination rate in squamate reptiles, our data suggest that the rate of ectopic recombination might be higher in the anole (and birds) than in mammals. It has been previously suggested that one of the conditions that makes mammalian genomes permissive to the amplification of L1 is a low rate of ectopic recombination (Eickbush and Furano 2002). The decay of L1 in mammals results from the accumulation of neutral mutation and small indels and not from the loss of element ends, as observed in anole and chicken. This difference in the decay of copies underlies a fundamental difference in the frequency of ectopic exchange between mammals and nonmammalian vertebrates and provides an explanation for the contrasted diversity and abundance of non-LTR retrotransposons among vertebrates.

#### Supplementary Material

Supplementary figure 1 is available at *Molecular Biology and Evolution* online (http://www.mbe. oxfordjournals.org/).

#### 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 *Anolis carolinensis* available. We also would like to thank Rochelle Clinton for her technical support. This research was supported by PSC-CUNY grant 61542-00-39 to Stéphane Boissinot.

### Literature Cited

- Abrusan G, Krambeck HJ, Junier T, Giordano J, Warburton PE. 2008. Biased distributions and decay of long interspersed nuclear elements in the chicken genome. Genetics. 178:573–581.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.
- Basta HA, Buzak AJ, McClure MA. 2007. Identification of novel retroid agents in *Danio rerio*, *Oryzias latipes*, *Gasterosteus* aculeatus and *Tetraodon nigroviridis*. Evol Bioinform. 3:179–195.
- Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature. 441:87–90.
- Ben-Aroya S, Mieczkowski PA, Petes TD, Kupiec M. 2004. The compact chromatin structure of a Ty repeated sequence suppresses recombination hotspot activity in *Saccharomyces cerevisiae*. Mol Cell. 15:221–231.
- Boissinot S, Chevret P, Furano AV. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol Biol Evol. 17:915–928.
- Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. 2006. Fitness cost of LINE-1 (L1) activity in humans. Proc Natl Acad Sci USA. 103:9590–9594.
- Boissinot S, Entezam A, Furano AV. 2001. Selection against deleterious LINE-1-containing loci in the human lineage. Mol Biol Evol. 18:926–935.
- Burke WD, Malik HS, Jones JP, Eickbush TH. 1999. The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol Biol Evol. 16: 502–511.
- Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. Genet Res. 42:1–27.
- Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. Annu Rev Genet. 23:251–287.
- Cooper DM, Schimenti KJ, Schimenti JC. 1998. Factors affecting ectopic gene conversion in mice. Mam Genome. 9: 355–360.
- Cost GJ, Feng Q, Jacquier A, Boeke JD. 2002. Human L1 element target-primed reverse transcription in vitro. EMBO J. 21:5899–5910.
- Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet. 35:41–48.
- Dewannieux M, Heidmann T. 2005. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. J Mol Biol. 349:241–247.
- 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.
- Eickbush TH, Furano AV. 2002. Fruit flies and humans respond differently to retrotransposons. Curr Opin Genet Dev. 12:669–674.
- Feng Q, Moran JV, Kazazian HH Jr., Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell. 87:905–916.

- Furano AV. 2000. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol. 64:255–294.
- Furano AV, Duvernell DD, Boissinot S. 2004. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. Trends Genet. 20:9–14.
- Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, Pollock DD, Jurka J. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. Genome Res. 17:992–1004.
- Gilbert N, Labuda D. 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. J Mol Biol. 298:365–377.
- Gish W. 1996–2004. WU-BLAST. Available from: http:// blast.wustl.edu/ (accessed May 18, 2009).
- Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33:W557–W559.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symp Ser. 41:95–98.
- Hull R. 2001. Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. International Committee on Taxonomy of Viruses. Arch Virol. 146:2255–2261.
- Karolchik D, Baertsch R, Diekhans M, et al. (12 co-authors). 2003. The UCSC genome browser database. Nucleic Acids Res. 31:51–54.
- 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.
- Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science. 256:1783–1790.
- Kordis D, Lovsin N, Gubensek F. 2006. Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. Syst Biol. 55:886–901.
- Lander ESLM, Linton B, Birren C, et al. (254 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature. 409:860–921.
- Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of unequal exchange in the containment of transposable element copy number. Genet Res. 52:223–235.
- Luan DD, Eickbush TH. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol Cell Biol. 15:3882–3891.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 72:595–605.
- Makalowski W. 2000. Genomic scrap yard: how genomes utilize all that junk. Gene. 259:61–67.
- Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol. 16:793–805.
- Marchler-Bauer A, Anderson JB, Derbyshire MK, et al. (25 coauthors). 2007. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. 35:D237–D240.
- Martin F, Maranon C, Olivares M, Alonso C, Lopez MC. 1995. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. J Mol Biol. 247:49–59.
- McClure MA. 1991. Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol Biol Evol. 8:835–856.

- McClure MA, Johnson MS, Feng DF, Doolittle RF. 1988. Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. Proc Natl Acad Sci USA. 85:2469–2473.
- McClure MA, Richardson HS, Clinton RA, Hepp CM, Crowther BA, Donaldson EF. 2005. Automated characterization of potentially active retroid agents in the human genome. Genomics. 85:512–523.
- Mikkelsen TS, Wakefield MJ, Aken B, et al. (62 co-authors). 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature. 447:167–177.
- Neafsey DE, Blumenstiel JP, Hartl DL. 2004. Different regulatory mechanisms underlie similar transposable element profiles in pufferfish and fruitflies. Mol Biol Evol. 21:2310–2318.
- Nekrutenko A, Li WH. 2001. Transposable elements are found in a large number of human protein-coding genes. Trends Genet. 17:619–621.
- Pascale E, Liu C, Valle E, Usdin K, Furano AV. 1993. The evolution of long interspersed repeated DNA (L1, LINE 1) as revealed by the analysis of an ancient rodent L1 DNA family. J Mol Evol. 36:9–20.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.
- 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.
- Skowronski J, Fanning TG, Singer MF. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. Mol Cell Biol. 8:1385–1397.
- Smit AFA, Hubley R, Green P. 2004. RepeatMasker Open-3.0. Institute of Systems Biology. Available from: http://repeatmasker.org/ (accessed May 18, 2009).

- Song M, Boissinot S. 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. Gene. 390:206–213.
- 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.
- Volff JN, Bouneau L, Ozouf-Costaz C, Fischer C. 2003. Diversity of retrotransposable elements in compact pufferfish genomes. Trends Genet. 19:674–678.
- Volff JN, Korting C, Schartl M. 2000. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. Mol Biol Evol. 17:1673–1684.
- Voliva CF, Jahn CL, Comer MB, Hutchison CA 3rd, Edgell MH. 1983. The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. Nucleic Acids Res. 11:8847–8859.
- Warren WC, Hillier LW, Marshall Graves JA, et al. (104 coauthors). 2008. Genome analysis of the platypus reveals unique signatures of evolution. Nature. 453:175–183.
- 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.
- Xia X, Xie Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. J Hered. 92:371–373.
- Zupunski V, Gubensek F, Kordis D. 2001. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. Mol Biol Evol. 18:1849–1863.

Scott Edwards, Associate Editor

Accepted April 18, 2009