

Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*

David A. Ray,^{1,5,6} Cedric Feschotte,^{2,5} Heidi J.T. Pagan,¹ Jeremy D. Smith,¹
Ellen J. Pritham,² Peter Arensburger,³ Peter W. Atkinson,³ and Nancy L. Craig⁴

¹Department of Biology, West Virginia University, Morgantown, West Virginia 26506, USA; ²Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA; ³Department of Entomology and Institute for Integrative Genome Biology, University of California, Riverside, California 92521, USA; ⁴Howard Hughes Medical Institute, Department of Molecular Biology & Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

DNA transposons, or class 2 transposable elements, have successfully propagated in a wide variety of genomes. However, it is widely believed that DNA transposon activity has ceased in mammalian genomes for at least the last 40 million years. We recently reported evidence for the relatively recent activity of *hAT* and *Helitron* elements, two distinct groups of DNA transposons, in the lineage of the vespertilionid bat *Myotis lucifugus*. Here, we describe seven additional families that have also been recently active in the bat lineage. Early vespertilionid genome evolution was dominated by the activity of *Helitrons*, *mariner*-like and *Tc2*-like elements. This was followed by the colonization of *Tc1*-like elements, and by a more recent explosion of *hAT*-like elements. Finally, and most recently, *piggyBac*-like elements have amplified within the *Myotis* genome and our results indicate that one of these families is probably still expanding in natural populations. Together, these data suggest that there has been tremendous recent activity of various DNA transposons in the bat lineage that far exceeds those previously reported for any mammalian lineage. The diverse and recent populations of DNA transposons in genus *Myotis* will provide an unprecedented opportunity to study the impact of this class of elements on mammalian genome evolution and to better understand what makes some species more susceptible to invasion by genomic parasites than others.

[Supplemental material is available online at www.genome.org.]

Transposable elements represent a substantial fraction of many eukaryotic genomes. For example, ~50% of the human genome is derived from transposable element sequences (Lander et al. 2001). Other genomes, especially those of plants, may consist of substantially higher proportions of transposable element-derived DNA (SanMiguel and Bennetzen 1998; Bennetzen 2000). Transposable elements are typically divided into two classes (Wicker et al. 2007). Class 1 is represented by the retrotransposons (LINEs, SINEs, LTRs, and ERVs). Class 2 includes the “cut-and-paste” DNA transposons, which are characterized by terminal inverted repeats (TIRs) and are mobilized by an element-encoded transposase. Currently, 10 superfamilies of cut-and-paste DNA transposons are recognized in eukaryotes (Feschotte and Pritham 2007).

While class 2 elements are widespread and active in a variety of eukaryotes, they have been thought to be transpositionally inactive in mammalian genomes. This conclusion was based primarily on the initial analyses of the human and mouse genome sequences. While both species harbor a significant number and a diverse assortment of DNA transposons, they show no signs of recent activity (Lander et al. 2001; Waterston et al. 2002). For example, there are more than 300,000 DNA elements recognizable in the human genome, which are grouped into 120 families and belong to five superfamilies. A large subset of these elements (40 families; ~98,000 copies) were integrated in the last 40–80 million years (Myr), but there remains no evidence for any hu-

man DNA transposon families younger than ~37 Myr (Pace and Feschotte 2007).

In stark contrast to the situation in human and mouse, there is mounting evidence for recent and substantial DNA transposon activity in the vespertilionid bat *Myotis lucifugus*. In one study, this conclusion was inferred from the presence of minimally diverged nonautonomous *hAT* transposons, the polymorphic status of some of these transposons in natural *Myotis* populations, and the discovery of an apparently full-length and potentially functional autonomous *hAT* transposon (Ray et al. 2007). In a separate report, an extensive and recently amplified population of *Helitrons*, a distinct subclass of DNA transposons, was discovered in the *M. lucifugus* genome (Pritham and Feschotte 2007). These surprising observations raised a number of intriguing questions about the activity of DNA transposons and its genomic impact in *Myotis* and other bats. In particular, we wondered whether the discovery of relatively recent *hAT* and *Helitron* activity in *Myotis* were merely coincidental or whether it reflected some unique biological or genomic features of bats in general or vespertilionids in particular that has allowed DNA transposons to invade and expand in these genomes, while such occurrences appear to be rare or nonexistent in other mammalian genomes. As a prelude to addressing these questions we examined *M. lucifugus* for additional traces of DNA transposon activity. Consequently, we explored the diversity, abundance, and evolutionary history of DNA transposons in the *M. lucifugus* genome.

Herein, we characterize seven new families of DNA transposons from the current whole-genome sequence of *M. lucifugus*. In addition, we provide additional information on the previously described *Myotis_hAT* transposons (Ray et al. 2007). Together, these elements represent three distinct superfamilies and exhibit

⁵These authors contributed equally to this work.

⁶Corresponding author.

E-mail David.Ray@mail.wvu.edu; fax (304) 293-6363.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.071886.107>.

clear signs of recent activity in the vespertilionid lineage, that is, within the last 40–50 Myr. We also identified a ninth transposon lineage that likely expanded prior to the chiropteran divergence. One family represents the youngest DNA transposon family so far recorded in any mammalian species, one that is likely still expanding in the genome. The discovery of this unprecedented level of DNA transposon activity in a mammalian genome represents a dramatic shift in our view of mobile element biology in mammals.

Results

hAT superfamily

The consensus sequence for *Myotis_hAT1* was previously described (Ray et al. 2007). To summarize, the entire sequence spans 2921 nucleotides with a single ORF consisting of bases 700–2631 that encodes an apparently intact transposase of 643 amino acids (aa). The present analysis confirms the characteristic target-site duplications (TSDs) for *Myotis hAT* elements, 8 bp with a central TA dinucleotide, and the typical short TIR (terminal inverted repeats) sequence. *Myotis_hAT1* and its nonautonomous derivatives are by far the most abundant family of elements analyzed in this study, with >96,000 hits of ≥ 100 bp in the current WGS data.

We identified a second family of *hAT* elements, called *hAT2_ML*. The TIRs of *hAT2_ML* are highly similar to *Myotis_hAT1*, but the internal sequence is only weakly similar. *hAT2_ML* is predicted to encode a 428-aa transposase with only minimal sequence similarity and 33% amino acid identity to the *Myotis_hAT1* transposase. Phylogenetic comparison to other *hAT* transposases confirms that *Myotis_hAT1* and *hAT2_ML* belong to the *hAT* superfamily and are closely related to *hAT1_MD* (Rebase Update, volume 12, issue 10), a *hAT* transposon family recently identified in the opossum *Monodelphis domestica* (Fig. 1A). RepeatMasker results using the full consensus as a query suggests that *hAT2_ML* is not nearly as widespread as *Myotis_hAT1* with only 2004 hits of ≥ 100 bp.

We also recovered a third *hAT* family member, *hAT3_ML*, from the *M. lucifugus* genome. The consensus sequence contains a large ORF spanning base pairs 1595–3403 and encoding a putative protein of 602 aa, which nests in phylogenetic analysis within a clade encompassing several mammalian *hAT* transposases (Fig. 1A). With its 3904-bp consensus, it is the longest of all the elements described in this study and is second to *Myotis_hAT1* in terms of its abundance with >12,000 elements of ≥ 100 bp identified in the current WGS assembly of *M. lucifugus*.

Tc1/mariner superfamily

Eukaryotic *Tc1/mariner* elements are divided into several anciently diverged lineages (Robertson 2002; Feschotte and Pritham 2007). Three distinct *Tc1/mariner* lineages have been previously identified in mammals and characterized most extensively in human: *pogo*, *Tc2*, and D34D *mariner*. We identified members of all three groups in *M. lucifugus* as well as the first mammalian member of the *Tc1* group. The *pogo* lineage is represented by *Tigger* elements in mammals. It is known to include both eutherian-wide and primate-specific families (Robertson 1996; Smit and Riggs 1996; Pace and Feschotte 2007). Thus, the presence of *Tigger* elements was somewhat expected in *M. lucifugus*. Indeed, we identified remnants of what appears to be an ancient family of

Tigger-like elements called *Tigger1_ML*. A tentative consensus was reconstructed from the alignment of the seven longest copies. The size (2.8 kb) and TIRs are consistent with other *Tigger* elements and the consensus is similar to consensus sequences in Repbase as *Tigger1a_CAR* and *Tigger1a_ART*, which represent *Tigger1*-like families specific to carnivores and artiodactyls, respectively. We also identified several highly eroded copies of a related family of *Tigger1* transposons in the European hedgehog, *Erinaceus europaeus* (e.g., accession AANN01830819, position 1599–4234). Given that hedgehog, carnivores, artiodactyls, and chiropterans are all part of the well-supported Laurasiatheria superorder, we hypothesize that *Tigger1_ML* amplified during the early part of the Laurasiatherian radiation. This is by far the oldest family of DNA transposons that we identified and, as it does not appear to be unique to chiropterans, we have chosen not to focus on the details of this family in the current study.

The *Tc2* group is sister to the *pogo* lineage and was first identified in *Caenorhabditis elegans* (Ruvolo et al. 1992). There are relatively few *Tc2*-like elements in human, and their amplification predates the split of eutherian mammals (Pace and Feschotte 2007). In contrast, we identified a distinct and much more recent family of *Tc2*-like elements in *M. lucifugus*. The *Tc2_1_ML* consensus is 1728 bp with 23 bp TIRs similar to known or newly identified *Tc2* elements. It contains a single ORF spanning positions 213–1559 and encoding a putative transposase of 448 aa. Phylogenetic analysis confirms that *Tc2_1_ML* belongs to the *Tc2* group, but does not cluster with either of the mammalian *Tc2*-like transposases identified in human (Kanga), opossum (*Tc2_MD2*), and tenrec (*Tc2_Et*), or with the human transposase-derived proteins POGK and POGZ (Fig. 1B). Thus, *Tc2_1_ML* defines a distinct lineage of *Tc2*-like elements. *Tc2_1_ML* is the least abundant of the DNA transposon families described in this study with only 589 hits ≥ 100 bp.

Two distinct lineages of canonical D34D *mariner* are known to occur in mammals: the *cecropia* subfamily, represented by *Hsmar1* in anthropoid primates and *Mmar1* in mouse, and the *irritans* subfamily, represented by *Hsmar2* in primates (Robertson 2002). We detected an abundant (2036 hits >100 bp) family of transposons in *M. lucifugus*, *Mlmar1*, that displays all the features of a canonical D34D *mariner*. The *Mlmar1* consensus is 1287 bp with 30-bp TIRs and a single ORF, spanning position 174–1184 and predicted to encode a 336-aa transposase. Surprisingly, phylogenetic analyses robustly place the predicted *Mlmar1* transposase within the *mauritiana* subfamily of *mariner* (Fig. 1C), a subfamily typified by the original *mos1* element active in *Drosophila mauritiana* (Hartl et al. 1997; Robertson 2002). We could not detect any members of the *irritans* or *cecropia* subfamilies of *mariner* in *M. lucifugus* using protein queries representing these lineages, although these two lineages have been identified in a broad range of mammals (Demattei et al. 2000; Robertson 2002; Waterston et al. 2002; Sinzelle et al. 2006), including other bat species (see below).

Tc1-like elements represent another distinct lineage of *Tc1/mariner* elements that are widespread and common in invertebrates and lower vertebrates (fish, frogs), but no *Tc1*-like element has previously been identified in mammals (Avancini et al. 1996; Leaver 2001; Sinzelle et al. 2005). Consistent with this observation, TBLASTN searches with representative *Tc1*-like transposases from fish or nematode return no significant hits in the complete genome of human, mouse, rat, and dog. In contrast, the same query returns hundreds of hits from the *M. lucifugus* WGS database. A single *Tc1*-like family, *Tc1_1_ML*, predominates in the

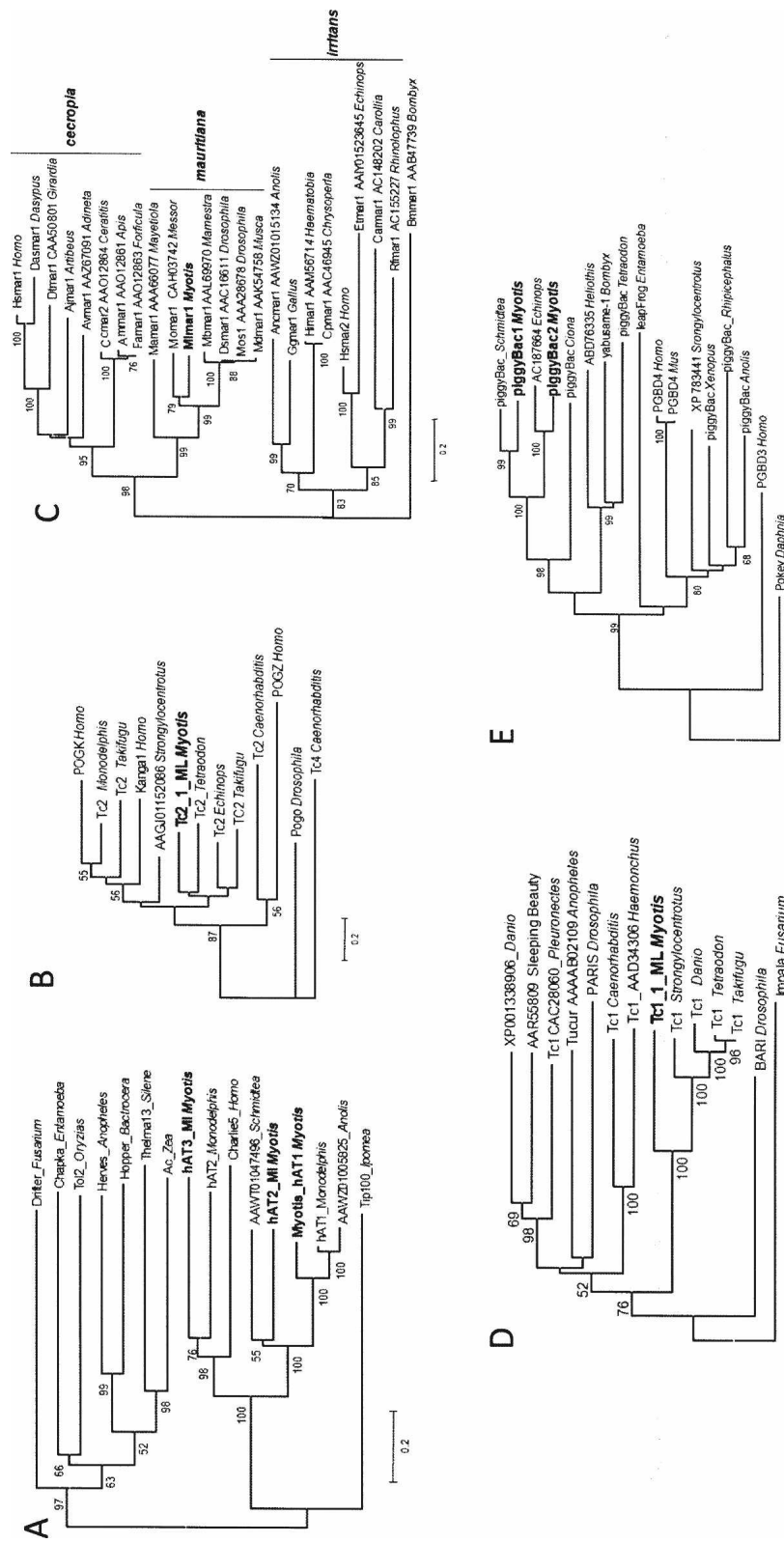


Figure 1. Results of phylogenetic analyses. (A) hAT transposases, (B) Tc2 transposases, (C) mariner transposases, (D) Tc1 transposases, (E) piggyBac transposases.

genome. The *Tc1_1_ML* consensus spans 1222 bp with 29 bp TIR with a 5'-CACTG-3' terminal motif typical of many *Tc1*-like elements. The ORF occupies position 150–1190 of the consensus and encodes a putative transposase of 346 residues. Phylogenetic analyses show that the *Tc1_1_ML* transposase forms a well-supported clade with *Tc1*-like transposases from fish (Fig. 1D). We detected 3788 *Tc1_1_ML* fragments of 100 bp or larger in the *M. lucifugus* WGS database. Many of the *Tc1_1_ML* elements are inserted in (TA) dinucleotide repeats, an insertion preference also observed for other *Tc1* elements (Vigdal et al. 2002). Of note is one subfamily derived from *Tc1_1_ML*. The consensus of this subfamily has an intact ORF that represents 2/3 (708 bp; 235 aa) of the complete ORF for *Tc1_1_ML*. Five hundred copies of this subfamily of *Tc1* are distributed throughout the *M. lucifugus* genome.

piggyBac superfamily

piggyBac-like elements have been identified in a wide range of animal species and *Entamoeba* (Sarkar et al. 2003; Pritham et al. 2005). Among mammals, *piggyBac* elements have so far only been characterized in the human genome, where they are predominantly represented by two families of nonautonomous elements (*MER85* and *MER75*) and by several stationary “domesticated” transposases (*PGBD1-5* genes) (Sarkar et al. 2003). We identified two families of *piggyBac*-like elements in *M. lucifugus*. *piggyBac1_ML* is defined by a 2626-bp consensus with short TIRs (15 bp) that are very similar to other *piggyBac* transposons. The consensus contains a 1719-bp ORF (position 587–2305) that likely encodes a transposase of 572 residues. Two potentially active elements, with intact ORFs and identifiable TSDs, were located in the available genome data. *piggyBac2_ML* is similar in length at 2639 bp and with a 583-aa transposase (nt 716–2467). Sequence comparison with other *piggyBac* transposases and phylogenetic analyses indicate that the two families are relatively closely related and form a strong cluster with the sea squirt, *Ciona intestinalis* (Fig. 1E). *Myotis piggyBac*-like elements tend to be flanked by the canonical 5'-TTAA-3' TSD. *piggyBac1_ML* elements number at least 2056 in the *M. lucifugus* genomes and are likely still mobilizing (see below). However, *piggyBac2_ML* elements are currently more numerous, with at least 3869 instances.

In summary, we identified a diverse array of *hAT*, *Tc1/mariner*, and *piggyBac* transposons in the *M. lucifugus* genome. The *Myotis hAT1* and *hAT3_ML* elements are by far the most abundant. As with *Myotis hAT1* elements (Ray et al. 2007), all of the newly discovered trans-

poson families are associated with multiple nonautonomous MITE subfamilies. It is beyond the scope of the present study to describe all of these subfamilies in detail, but we note that nonautonomous elements appear to have outnumbered their autonomous partners, as is typically observed for DNA transposon families in humans and most other eukaryotes. (Feschotte et al. 2002; Pace and Feschotte 2007)

Taxonomic distribution of DNA transposons identified in *Myotis lucifugus*

PCR-based analyses of 15 taxa representing nine bat families were performed using two oligonucleotide primer pairs for each transposon family (Fig. 2). Results indicate that the families described

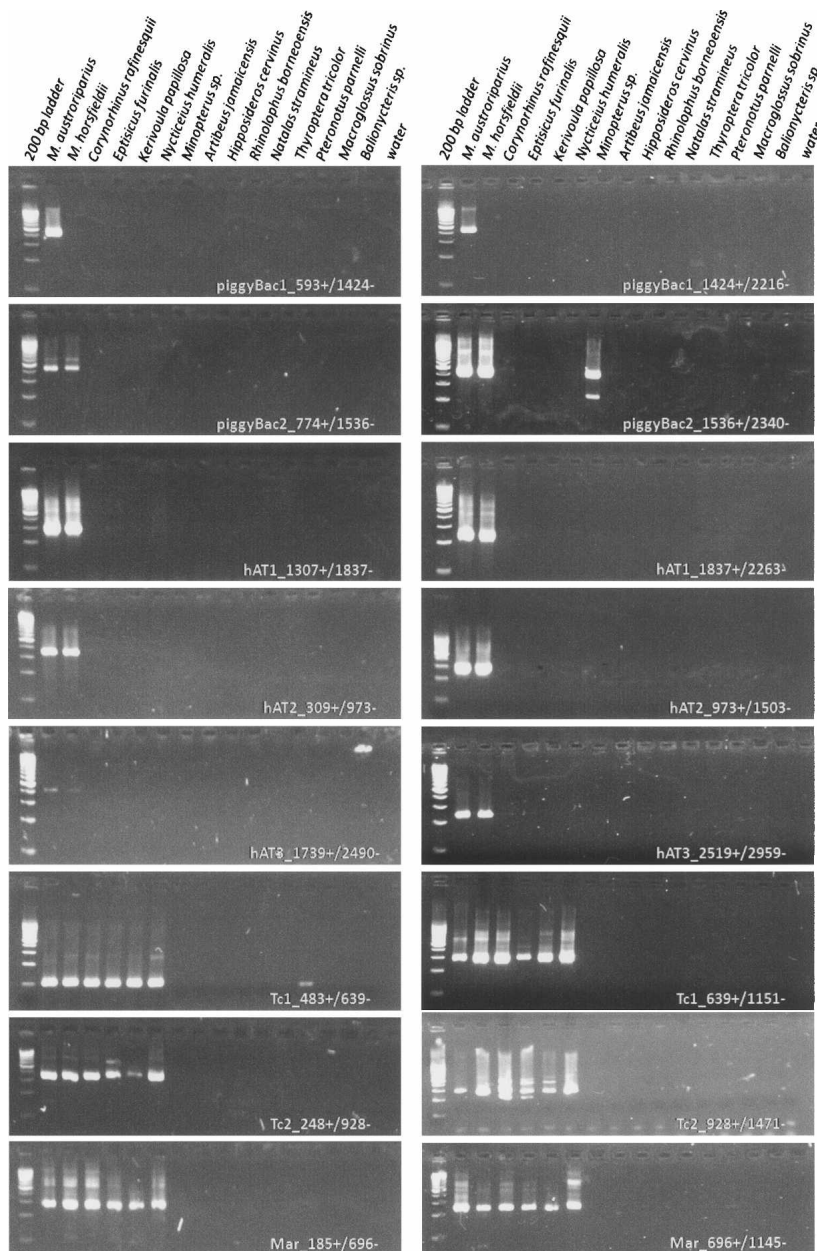


Figure 2. Electropherogram results from PCR analysis of a taxonomically diverse panel of bat genomes. Oligonucleotide primer pairs are indicated in white.

have limited taxonomic distributions. *Tc1_1_ML*, *Tc2_1_ML*, and *Mlmar1* appear to be limited to Vespertilionidae with the possible exception of *Tc1*-like sequences in *Pteronotus parnellii* (Family Mormoopidae). All three *hAT* families are restricted to the genus *Myotis*. *piggyBac2_ML* shares a similar distribution to the *hAT* elements, but positive results were obtained from the lone representative of Miniopteridae, a family that was recently elevated from subfamily status within Vespertilionidae (Miller-Butterworth et al. 2007). Amplification of the miniopterid and mormoopid representatives suggests the need for additional exploration of these taxa. PCR data using *piggyBac1_ML*-derived oligonucleotides suggests that these elements are even further restricted. Amplification was obtained in *Myotis austroriparius*, a fellow North American bat, but not the Asian representative, *Myotis horsfieldii*. We acknowledge that these results do not confirm that no class 2 transposons are active in other bat lineages. Instead, they lend support to the hypothesis that these particular transposons have a limited taxonomic distribution.

With the exception of *Tigger1_ML*, whose amplification likely predates the split of Chiroptera or perhaps even Laurasiatheria, TBLASTN and BLASTN queries of chiropteran taxa with the consensus transposon sequences from *M. lucifugus* yielded only very few significant hits in other bat species. A handful of hits were obtained with the *Mlmar1* predicted transposase in the phyllostomid bats *Artibeus jamaicensis* and *Carollia perspicillata*, and in the rhinolophid bat, *Rhinolophus ferrumequinum*. Interestingly, the hits correspond to *cecropia*-like and the previously described *irritans*-like *mariners* (Sinzelle et al. 2006) (Fig. 1C). Thus, it appears that distinct types of *mariners* have colonized different bat lineages. *Mlmar1* is most closely related to *Mbou-*

mar1, a *mauritiana*-like *mariner* from the ant *Messor bouvieri* (see also phylogeny in Fig. 1C). The transposase regions are 75% identical at the nucleotide level and 85% similar at the protein level. This level of sequence similarity is remarkable given the evolutionary distance of their host species, >800 Myr (Gu 1998; Blair and Hedges 2005). We could not detect any traces of other *mauritiana*-like *mariner* transposases in any of the ~30 additional mammalian species nor any other chordate species represented in the NCBI databases.

Age estimations of DNA transposon families

Table 1 and Figure 3 show the estimated ages for each family based on average divergence from the consensus sequence and an estimated neutral mutation rate of -2.366×10^{-9} (see Methods). Age analyses of the eight youngest families produce a clear temporal pattern indicative of successive expansions in the genome. The earliest invasions began with *Tc2*-like and *mariner*-like elements and were followed by the colonization of *Tc1_1_ML*, *hAT*-like, and *piggyBac*-like families.

In mammals, CpG dinucleotides within repeats are generally heavily methylated, and thus tend to degrade into TpG or CpA dinucleotides at a substantially higher rate than other base combinations (Coulondre et al. 1978; Razin and Riggs 1980; Xing et al. 2004). Thus, lower divergence estimates were expected and observed when calculated using sequences from which the CpG dinucleotides had been removed. CpG to non-CpG mutations densities were calculated for all families (Table 2). The youngest subfamilies, derived from *piggyBac1_ML*, produced consistently lower ratios than any other subfamily. This is most likely ex-

Table 1. Divergence values and estimated activity periods for the DNA transposon groups described

Element family	CpG Sites Included					CpG Sites Removed					
	N ^a	Mean distance	Substitution model	Std. error	Range	Age estimate ^b (Myr)	Mean distance	Model	Std. error	Range	Age estimate ^b (Myr)
<i>npiggy1_156</i>	181	0.003	K80	0.0004	0.0000–0.0193	1.1–1.4	0.0021	K80	0.0003	0.0000–0.0193	0.76–1.0
<i>piggyBac1_ORF</i>	15	0.0046	K81uf	0.001	0.0012–0.0145	1.5–2.4	0.0038	F81	0.0011	0.0006–0.0148	1.1–2.1
<i>npiggy1_239</i>	153	0.0055	HKY+G	0.0004	0.0000–0.0311	2.2–2.5	0.0029	F81+G	0.0004	0.0000–0.0220	1.1–1.4
<i>nhAT3_200</i>	200	0.0155	HKY+G	0.0007	0.0000–0.0579	6.3–6.9	0.0079	HKY+G	0.0006	0.0000–0.0436	3.1–3.6
<i>piggyBac2</i>	5	0.0159	HKY	0.0013	0.0121–0.0197	6.2–7.3	0.0100	HKY	0.0008	0.0074–0.0119	3.9–4.6
<i>hAT3_ORF</i>	13	0.0177	HKY+G	0.0013	0.0074–0.0233	6.9–8.0	0.0084	HKY+G	0.0012	0.0040–0.0184	3.0–4.1
<i>nhAT2_730</i>	460	0.0207	TVM+G	0.0003	0.0073–0.0484	8.6–8.9	0.0142	TVM+G	0.0003	0.0014–0.0449	5.9–6.1
<i>hAT1_ORF</i>	42	0.0209	GTR+G	0.0008	0.0132–0.0393	8.5–9.2	0.0149	TVM+G	0.0007	0.0087–0.0280	6.0–6.6
<i>npiggy2_345</i>	92	0.0211	TVM+G	0.001	0.0000–0.0599	8.5–9.3	0.0125	K81uf	0.0008	0.0000–0.0367	5.0–5.6
<i>hAT2_ORF</i>	7	0.0213	HKY+G	0.0019	0.0155–0.0290	8.2–9.8	0.0154	HKY	0.0013	0.0110–0.0205	6.0–7.1
<i>nhAT2_1124</i>	47	0.0235	TVM+G	0.0015	0.0111–0.0717	9.3–10.6	0.0161	TVM+G	0.0012	0.0055–0.0516	6.3–7.3
<i>nhAT2_525</i>	138	0.0246	TVM+G	0.0009	0.0069–0.0520	10.0–10.8	0.0164	K81uf+G	0.0007	0.0020–0.0430	6.6–7.2
<i>nhAT1_186</i>	32	0.0296	K81uf+G	0.0025	0.0096–0.0707	11.5–13.6	0.0242	K81uf+G	0.0026	0.0049–0.0689	9.1–11.3
<i>nhAT1_239a</i>	45	0.0304	K81uf+G	0.0025	0.0086–0.0706	11.8–13.9	0.0215	K81uf	0.0022	0.0000–0.0591	8.2–10.0
<i>nTc1_452</i>	36	0.0346	HKY+G	0.0019	0.0147–0.0744	13.8–15.4	0.0201	HKY	0.0013	0.0077–0.0481	8.0–9.0
<i>nTc1_950</i>	41	0.0365	HKY+G	0.0015	0.0218–0.0593	14.8–16.1	0.0225	HKY	0.0011	0.0122–0.0354	9.0–10.0
<i>nMar_1285</i>	69	0.0449	TVM+G	0.001	0.0296–0.0695	18.6–19.4	0.0283	TVM+G	0.0008	0.0136–0.0454	11.6–12.3
<i>nMar_1265</i>	37	0.0473	TVM+G	0.0017	0.0310–0.0716	19.3–20.7	0.0300	TVM+G	0.0013	0.0181–0.0503	12.1–13.2
<i>Tc2_ORF</i>	6	0.0495	TrN+G	0.0017	0.0452–0.0543	20.2–21.6	0.0397	TrN	0.0017	0.0355–0.0462	16.1–17.5
<i>Mar_ORF</i>	113	0.0566	GTR+G	0.0011	0.0330–0.1046	23.5–24.4	0.0324	GTR+G	0.0008	0.0186–0.0751	13.4–14.0
<i>nTc2_555</i>	53	0.0694	TVM+G	0.0015	0.0464–0.0993	28.7–30.0	0.0425	TVM+G	0.0013	0.0219–0.0638	17.4–18.5
<i>nTc2_527</i>	25	0.0773	GTR+G	0.0026	0.0504–0.1005	31.6–33.8	0.0422	TIM	0.0022	0.0249–0.0672	16.9–18.8
<i>Helibat</i>	—	—	—	—	—	30–36 ^c	—	—	—	—	—

Table is arranged in increasing divergence estimates. “n” in the name indicates internally deleted, nonautonomous variants of the presumed autonomous element.

^aSample size for distance calculation.

^bBased on estimated mutation rate of 2.366×10^{-9} .

^cFrom Pritham and Feschotte (2007).

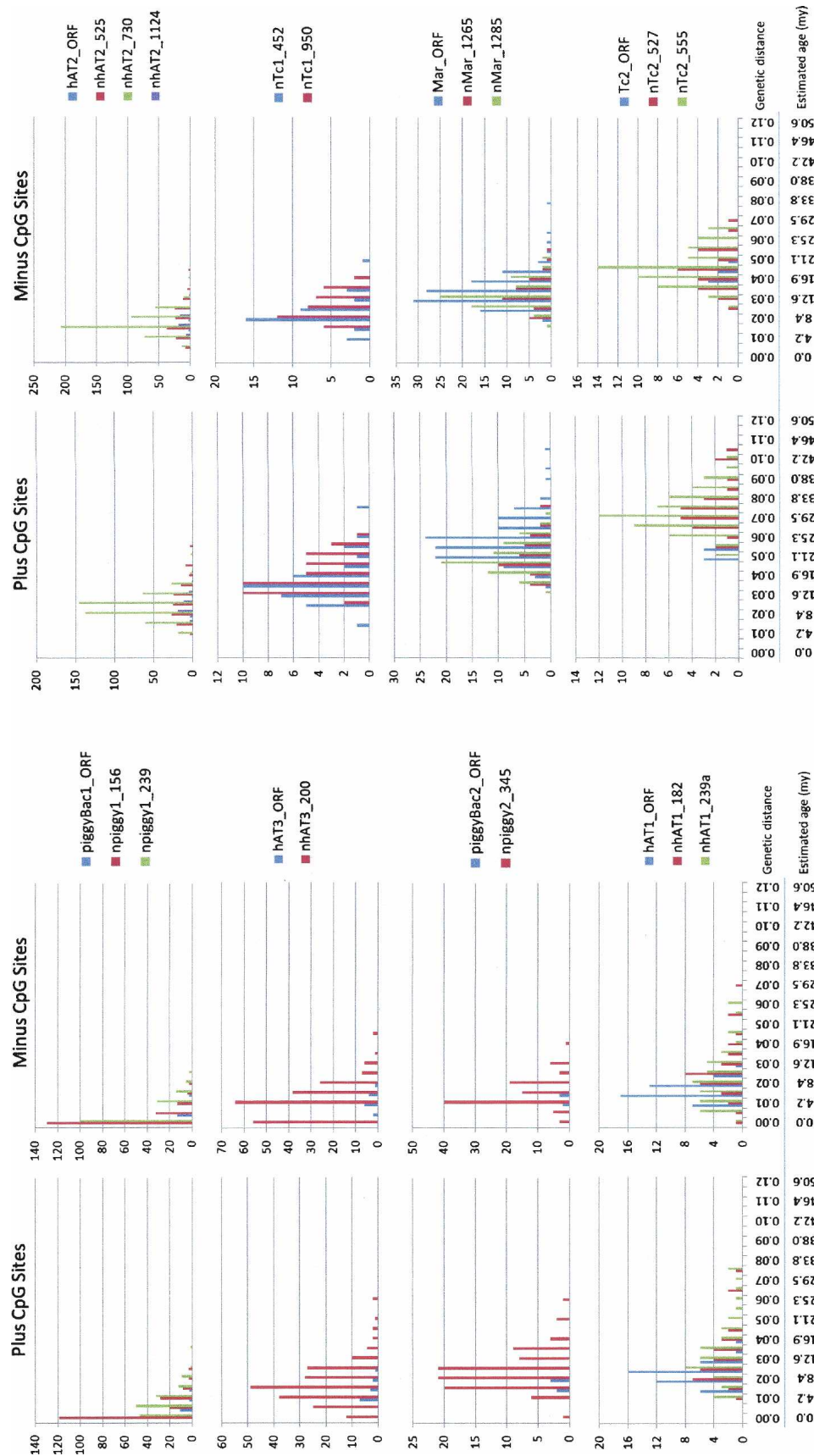


Figure 3. Divergence estimates and age ranges for each autonomous and nonautonomous transposon family described.

Table 2. CpG vs. nonCpG mutation ratios for each group of transposons described herein

Element	CpG:nonCpG mutation ratio
<i>piggyBac1_ML</i>	4.48
<i>npiggy1_156</i>	2.57
<i>npiggy1_239</i>	1.56
<i>piggyBac2_ML</i>	9.40
<i>npiggy2_345</i>	7.76
<i>Myotis_hAT1</i>	10.38
<i>nhAT1_186</i>	4.91
<i>nhAT1_239a</i>	4.55
<i>hAT2_ML</i>	8.83
<i>nhAT2_730</i>	10.21
<i>nhAT2_1124</i>	11.42
<i>nhAT2_525</i>	8.53
<i>hAT3_ML</i>	8.46
<i>nhAT3_200</i>	10.33
<i>nTc1_452</i>	10.06
<i>nTc1_950</i>	9.06
<i>Mar_ML</i>	8.16
<i>nMar_1265</i>	8.33
<i>nMar_1285</i>	8.24
<i>Tc2_ML</i>	8.46
<i>nTc2_527</i>	7.74
<i>nTc2_555</i>	7.92
Mean	7.79

The mean value was calculated using all estimations except those of *piggyBac1_ML* derived elements. See text for details. “n” in the name indicates internally deleted, nonautonomous variants of the presumed autonomous element.

plained by their very recent deposition in the genome (see below) and because they have not resided long enough to accumulate many mutations of any kind. This hypothesis, however, does not explain the relatively low ratios observed for the *nhAT1* subfamilies, which we believe were mobilized as much as 10 Myr earlier. Average CpG dinucleotide mutation densities were approximately eight times higher than non-CpG sites. We note that this is higher than the rate observed for primates ($\sim 6\times$; Xing et al. 2004), bringing to mind the hypothesis that increased methylation, and consequently, high CpG:non-CpG mutation ratios may serve as a genomic defense mechanism against large bursts of transposable element activity (Schmid 1998; Xing et al. 2004).

Current activity by *piggyBac1_ML*

Both genetic distance and CpG analyses above provide evidence for very recent and likely ongoing *piggyBac1_ML* activity. Indeed, a large fraction of the nonautonomous *piggyBac1* elements identified in *M. lucifugus* are strictly identical (65.7% of the elements in the *npiggy_156* subfamily and 30.7% in the *npiggy_239* subfamily), which strongly indicates that they have been inserted very recently. As one can observe in Figure 3, most of the DNA transposon families analyzed in this study appear to have reached an amplification peak and then entered a period of re-

duced activity and extinction. In contrast, the *npiggy_156* subfamily appears to be nearing its amplification peak, suggesting that *piggyBac1_ML* may still be active in the *Myotis* genome. The identification of two full-length elements with intact ORFs in the current genome sequence data further support this hypothesis. Further evidence for recent *piggyBac1_ML* activity is in the observation that we were able to experimentally isolate intact preintegration sites for *npiggy_156* elements in a related species, *M. austroriparius* (Fig. 4). Finally, experimental results from a diverse panel of bat species indicate that *piggyBac1_ML* is present in only a subset of species in the genus (Fig. 2). Thus, we can confidently state that this particular subfamily was mobilized within the last $\sim 8\text{--}12$ Myr (Stadelmann et al. 2007).

Nested insertion analysis

As a second, independent method to estimate relative ages of the elements, we compared the number and identity of insertions within individual transposon copies. Assuming that DNA transposon insertions occur fairly randomly throughout the genome, older elements should have higher proportions of nested insertions than younger elements. Furthermore, while it is possible for representatives from older transposon families to become nested within younger families if their periods of activity overlapped, these instances should be relatively rare. RepeatMasker hits for each family of transposons were subjected to a second RepeatMasker analysis using a custom library that included each of the new transposable elements described here and *Ves*, a SINE found in the suborder Yangochiroptera (Borodulina and Kramerov 1999, 2005). We limited ourselves to these elements because their sequences have been well characterized here or elsewhere (Borodulina and Kramerov 1999; Ray et al. 2007) and their target site duplications are easily identified.

Analysis of nested insertions supports the age hierarchy suggested by divergence estimates. Among the *hAT* and *piggyBac* elements examined, only two nested DNA transposon insertions were recovered—two *piggyBac1* insertions inside a *Myotis_hAT1* element (Table 3)—lending support to the hypothesis that these families are the most recently active DNA transposon families in the *M. lucifugus* genome. In contrast, the older elements represented by the *Tc1*, *Tc2*, and mariner families have been subjected to numerous interruptions by the more recently active elements. *Mlmar1* elements have suffered nested insertions from all other families except *Tc2_1_ML*, which is likely older. It is interesting that the *Tc2* elements have only accumulated two nested insertions by their fellow DNA transposons. This could be explained by the low occurrence of these elements ($n = 593$) and the relatively small size of these insertions (~ 560 bp). *Mariner* elements are present in larger numbers (>2000) and larger average size (778 bp), and thus present approximately five times the number of potential insertion targets during an only slightly shorter time frame. Analyses of *Ves* SINE nested insertions present a similar

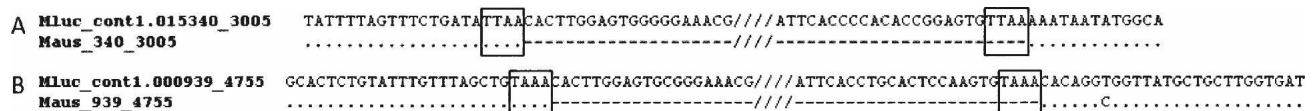


Figure 4. Filled and intact preintegration site sequences for individual nonautonomous *piggyBac1_ML* elements from *M. lucifugus* and *M. austroriparius*, respectively. Target site duplications for each insertion are shaded. (A) Locus 340_3005. (B) Locus primer sequences (5'–3') for the two loci presented are as follows: *Mluc*_cont1.015340_3005_L—TTGTACCAAAGGTGCCAAA, *Mluc*_cont1.015340_3005_R—TTTCCTCATATACCATCCATTTT; *Mluc*_cont1.000939_4755_L—CAAAATAAGGGAGAAAGGAAACA, *Mluc*_cont1.000939_4755_R—GGGCTGAGAAACAAGATCCA. (*Mluc*) *M. lucifugus*; (*Maus*) *M. austroriparius*.

Table 3. Results of nested insertion analysis for each DNA transposon family described herein

Element family	Total bp analyzed	Total bp verified nested insertions	Percent nested insertions	Insertion type	Quantity
<i>piggyBac1_ML</i>	620,634	214	0.03%	Ves	1
<i>hAT3_ML</i>	2,649,287	5,029	0.19%	Ves	24
<i>piggyBac2_ML</i>	1,495,500	6,829	0.46%	Ves	32
<i>Myotis_hAT1</i>	4,835,096	27,258	0.56%	<i>piggyBac1_ML</i>	2
				Ves	126
<i>hAT2_ML</i>	1,217,239	10,814	0.89%	Ves	52
<i>Tc1_1_ML</i>	1,750,122	9,619	0.55%	<i>Myotis_hAT1</i>	4
				<i>hAT3_ML</i>	2
				<i>piggyBac1_ML</i>	3
				Ves	36
<i>Mlmar1</i>	1,613,024	25,975	1.61%	<i>Myotis_hAT1</i>	3
				<i>hAT2_ML</i>	2
				<i>hAT3_ML</i>	3
				<i>piggyBac1_ML</i>	2
				<i>piggyBac2_ML</i>	1
				<i>Tc1_1_ML</i>	2
				Ves	98
<i>Tc2_ML</i>	335,994	5,964	1.78%	<i>piggyBac1_ML</i>	1
				<i>Tc1_1_ML</i>	1
				Ves	22

(Ves) VesSINE (Borodulina and Kramerov 1999).

picture. The four most recently active elements, *piggyBac1_ML*, *piggyBac2_ML*, *hAT2_ML*, and *hAT3_ML* have been subjected only to Ves insertions.

Nested self-insertions are theoretically possible and were expected when the analysis was performed. Indeed, there were some possible instances observed. However, in each of these cases the TSDs were not clearly identifiable. Because TSD identification was one a priori criterion for assessing the presence of any nested insertion (see Methods), we were forced to exclude these from Table 3.

Finally, one would expect older elements to have accumulated more nested insertions than more recently mobilized elements. To test this hypothesis, we examined the percent nested insertion content for all instances of each family of elements recovered from the *M. lucifugus* genome. Figure 5 illustrates a clear pattern of younger elements (i.e., *piggyBac*-like and *hAT*-like) having lower relative nested insertion content when compared with older elements from the *Tc1/mariner* superfamily. The

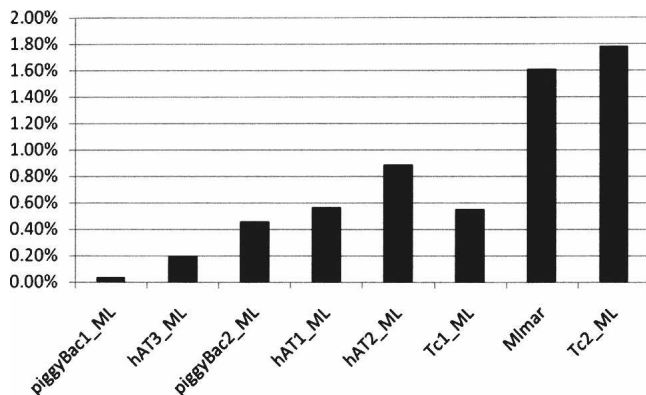


Figure 5. Proportion of insertions nested within instances of each DNA transposon family described. Families are arranged along the X-axis in order of age based on genetic divergence estimates in Table 3. Precise values for can be found in Table 3.

one exception to this pattern is *Tc1_1_ML*, which exhibits a relatively low percentage of nested insertions. It is unclear as to why this family exists as an outlier.

Discussion

DNA transposons are widespread and a diversity of elements have been recently active in the genomes of many eukaryotes, including lower vertebrates (Aparicio et al. 2002; Koga et al. 2006). In contrast, initial analyses of the transposable element landscape in the complete human, mouse, rat, and dog genomes have led to the common belief that mammalian genomes are devoid of recently active DNA transposons (Lander et al. 2001; Waterston et al. 2002; Gibbs et al. 2004; Lindblad-Toh et al. 2005). Supporting this idea, a detailed investigation of the evolutionary history of human DNA transposons revealed that many transposons were intensively active during early primate evolution (~80–40 million years ago [Mya]), but this activity ceased approximately

~40 Mya (Pace and Feschotte 2007). Analyses of the rodent and carnivore lineages reveal a similar trend, with no DNA transposon family identified in either mouse or dog that is significantly younger than 50 Myr (Waterston et al. 2002; J. Pace and C. Feschotte, unpubl.).

We have identified seven new families of DNA transposons in the *M. lucifugus* genome which, in stark contrast to observations made for other mammalian genomes, have apparently been active within the last 40 Myr. Several lines of evidence support this conclusion. First, the elements appear to be almost exclusively limited to Vespertilionidae, and in some cases to selected taxa within the family. Second, the level of sequence divergence among copies (and to their respective consensus sequences) suggests that these families were active from ~36 Mya to the present. Analysis of nested insertions suggests a hierarchical pattern of insertion consistent with sequence divergence estimates. Third, two potentially full-length and intact *piggyBac1_ML* transposons could be identified in the available genome sequence data, supporting the hypothesis that the expansion of these elements and their nonautonomous relatives is ongoing. Finally, recent activity of *piggyBac1_ML* was verified by recovering “empty” orthologous sites in a closely related *Myotis* species and by the limited taxonomic distribution of *piggyBac1_ML* to North American *Myotis* species. Together with the recent study reporting the massive amplification of *HeliBats* 30–36 Mya in the lineage of *M. lucifugus* (Pritham and Feschotte 2007), our data demonstrate that the genomes of vespertilionid bats have been subjected to multiple waves of amplification of diverse DNA transposons, from ~40 Mya to the present. One lineage of *piggyBac*, *piggyBac1_ML*, is likely still active. To our knowledge, this is the first documented evidence of the recent activity of a diverse population of DNA transposons in any mammal.

At least 3.5% of the available genome data is made up of sequence derived from the transposons described herein. It should be noted, however, that this number is very likely an underestimate. The previous study of *hAT1* elements in *M. lucifugus* revealed an abundance of nonautonomous *nhAT* families,

which have similar TIRs, but little internal sequence similarity with each other or with larger autonomous elements (Ray et al. 2007). Because of this lack of similarity, RepeatMasker fails to identify a majority of the previously identified *nhAT* elements when the consensus for *Myotis_hAT1* is used as the part of the database. Indeed, masking the genome using the *Myotis_hAT1* sequence produces only ~21,000 hits greater than 100 bp. Creating a database that contains the full-length consensus and the *nhAT* elements described by Ray et al. (2007) as a single repeat database increases the number of hits to >96,000. We believe that the other transposon families described herein will follow a similar pattern, and therefore, our estimate of the amount of DNA derived from DNA transposons is almost certainly too low. Add the impact of *HeliBats*, which accounts for at least an additional 3% of the *M. lucifugus* genome (Pritham and Feschotte 2007), and the as yet uncharacterized influence of retrotransposons suggests an extremely dynamic genomic landscape over the past 40 Myr of vespertilionid bat evolution.

Myotis is one of the most species-rich of all mammalian genera, and repeated waves of transposon activity suggest a mechanism for generating the genetic variability necessary to produce its tremendous species diversity. Interestingly, a recent analysis of *Myotis* phylogeny based on nuclear and mitochondrial DNA sequences becomes more intriguing when one considers the data presented here. Stadelmann et al. (2007) found that a burst of *Myotis* diversification occurred ~12–13 Mya. These dates correspond well to the estimated time during which the most active DNA transposon families were expanding in the *Myotis* genome (Fig. 2).

The young age and narrow taxonomic distribution of all (but one) of the transposon families identified in *M. lucifugus* raises the puzzling question of their evolutionary origin. Based on the data presented here and elsewhere (Pace and Feschotte 2007; Ray et al. 2007), we currently favor the hypothesis that most of these families hail from repeated episodes of horizontal introduction of founder autonomous transposons from yet unknown source(s) occurring at different evolutionary time points in the lineage of *M. lucifugus*. The only alternative hypothesis would be that each family arose from active transposons that were vertically inherited during chordate evolution. We consider this hypothesis less likely, because the lack of closely related elements in any of the 30+ other mammalian species (including three nonvespertilionid bats) for which complete or large amounts of genome sequence data are currently available would entail that these elements have persisted in the lineage of vespertilionid bats, while being systematically lost in all other lineages. Under a scenario dominated by horizontal introduction, it appears that *mariner*, *Tc1*, and *Tc2*-like elements colonized the ancestral vespertilionid genome prior to diversification of the family. Later invasions of the *hAT*-like and *piggyBac*-like elements may have occurred prior to diversification of *Myotis* and, in the case of *piggyBac1_ML*, New World *Myotis*. Such a scenario of repeated horizontal transfer is consistent with the well-known propensity and the apparent necessity of class 2 elements to undergo horizontal transfers (Hartl et al. 1997; Robertson 2002). A more thorough analysis of the distribution of these elements in bats and other vertebrates will shed additional light onto their origin.

These data also raised the puzzling question as to why bats (and Vespertilionidae in particular) exhibit such a high level of recent transposon activity, while these types of elements have seemingly gone extinct in other mammalian lineages examined. While we are currently unable to answer this question, we note

that bats exhibit a variety of life history traits unique among mammals (flight, high population sizes and densities, torpor) that provide potential avenues for future research. One of the most intriguing of these characteristics is the fact that bats are notorious viral reservoirs (Calisher et al. 2006). Like viruses, DNA transposons are examples of genomic parasites, and the ability of bats to safely harbor large loads of a variety of viruses may suggest similar genetic tolerance with regard to DNA transposons. On the other hand, viruses remain the best candidates as potential vectors for the horizontal introduction of DNA transposons and other TEs (Miller and Miller 1982; Fraser et al. 1983; Jehle et al. 1998; Piskurek and Okada 2007). Thus, the propensity of bats to tolerate massive and diverse viral infections may have facilitated the recurrent horizontal introduction of DNA transposons and/or their evolutionary persistence in vespertilionid bats.

In conclusion, we have provided evidence that the previously identified trend toward DNA transposon extinction in mammals is not universal and that a wide diversity of DNA transposons have been active throughout the diversification of the vespertilionid bats, that is, within the last 40 Myr. Furthermore, we propose that the genus *Myotis* is an excellent candidate for studying the impact and influence of mobile elements, especially class 2 elements, on the evolution of genomic, species, and ecological diversity in mammals. Concomitantly, studies of bats may represent a unique opportunity to investigate life history characteristics that make some organisms more susceptible to transposon activity than others.

Methods

Transposon identification and consensus reconstruction and classification

We carried out initial computational searches from January through April 2007 against the *M. lucifugus* WGS data deposited at NCBI (<http://www.ncbi.nlm.nih.gov/>, GenBank accession no. AAPE000000000; K. Lindblad-Toh, pers. comm.). At that time, the WGS data set represented ~73% of the unassembled *M. lucifugus* genome with contig sizes averaging ~2.4 kb (Pritham and Feschotte 2007). The WGS data was queried using TBLASTN to detect the presence of coding sequences related to all known DNA transposon superfamilies (list of queries available upon request). When significant hits were obtained (generally E -value $\ll 10^{-5}$), complete sequence entries spanning the top 10–50 subject hits (depending on the overall abundance and sequence heterogeneity of the hits) were retrieved from the database and aligned using a local installation of MUSCLE (Edgar 2004) or CLUSTAL (Thompson et al. 1994). Sequences were trimmed to eliminate unalignable flanks and leave only TSDs, noncoding regions, and the presumed coding regions of the elements. Complete alignments used to determine the full-length consensus sequence for each family are presented as Supplemental Alignments 1–8. Consensus sequences were derived from each multiple alignment based on a majority rule. For each consensus, coding sequences were predicted using ORF Finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) and, if necessary, refined manually at ambiguous positions in the consensus using multiple alignments of protein sequences from individual copies. Unless previous nomenclature had been established, families and subfamilies were disclosed according to the standard principles of TE classification (Wicker et al. 2007).

hAT3_ML was discovered independently using methods described in Arensburger et al. (2005). After its initial characteriza-

tion, it was subjected to the same methods as all other elements described herein. All transposon consensus sequences used for this manuscript have been deposited in Repbase (<http://www.girinst.org/repbase/index.html>).

Age estimation of *M. lucifugus* transposons

We estimated ages for each transposon family by extracting full- or near full-length ORF's based on the coordinates of the repeats drawn from the output files of a locally implemented version of RepeatMasker 3.1.6. Additionally, we hypothesized that nonautonomous subfamilies were likely mobilized by autonomous elements sharing the same TIRs, and thus were deposited during the same time period as the autonomous elements. We therefore collected representatives from a limited number of nonautonomous derivatives for all families. All extracted elements were aligned with their respective consensus sequence using MUSCLE (Edgar 2004). The most appropriate model for estimating nucleotide divergence from each consensus was determined using MODELTEST v3.7 (Posada and Crandall 1998) and genetic distances were calculated in PAUP* v4.0b10 (Swofford 2002).

Specific neutral mutation rate estimates for bats are not currently available. Thus, we utilized published sequences from representative vespertilionid (*Nycticeinops schlieffeni*, *Myotis tricolor*, *Scotophilus dinganii*, *Eptesicus hottentotus*, *Cistugo lesueurii*, and *Cistugo seabrai*) and miniopterid (*Miniopterus natalensis*, *M. inflatus*, *M. fraterculus*, *M. australis*, and *M. macrocneme*) taxa to obtain an estimate. Specifically, we obtained sequence data generated by Eick et al. (2005) for four introns (AJ865401-5, AJ865438-43, AJ865646-50, AJ865683-88, AJ866297-301, AJ866329-34, AJ866349-53, AJ866383-87). We constructed a concatenated alignment using MUSCLE (Edgar 2004) and determined the most appropriate evolutionary model for the data using MODELTEST v3.7 (Posada and Crandall 1998), TIM+G ($\Gamma = 1.2597$). We then estimated the average genetic distance between families Miniopteridae and Vespertilionidae as defined by Eick et al. (2005) using PAUP* v4.0b10 (Swofford 2002). By incorporating the estimated divergence time between these two families, ~45 Mya (Eick et al. 2005; Miller-Butterworth et al. 2007), we arrived at an estimated mutation rate of 2.366×10^{-9} , which was then used to determine likely periods of activity. Representative alignments for several nonautonomous subfamilies are presented in Supplemental Figure S1. All alignments are available upon request.

We estimated ages for each subfamily of elements using sequences as recovered from the genome and sequences from which the CpG dinucleotides had been removed (see Table 1). We also estimated the relative ratio of CpG to non-CpG mutations for each family (Table 2). This was accomplished using a modified Perl script originally designed to estimate CpG dinucleotide mutation rates in primate *Alu* elements (Xing et al. 2004) and modified for this work.

Nested insertion analysis

We extracted all RepeatMasker hits for each family of transposons and subjected each to

a second RepeatMasker analysis using a custom library that included each of the new transposable elements described here and Ves. We validated each nested insertion manually by checking for the presence of the expected target-site duplications.

Identification of intact preintegration sites in *M. austroriparius*

To test the hypothesis that *piggyBac1_ML* has been recently active in *Myotis*, we designed oligonucleotide primers to survey the presence/absence of 15 individual insertions of a nonautonomous *piggyBac1_ML* subfamily (*npiggy_156*) in *M. austroriparius*, which diverged from *M. lucifugus* ~8–12 Mya (Stadelmann et al. 2007). Primers were designed in the genomic regions flanking the elements using sequence data from *M. lucifugus* and tested on a panel of 10 individuals from natural populations of *M. austroriparius*. Details on sample collection, DNA extraction, amplification, cloning, and sequencing are as previously described (Ray et al. 2007). Sequences from the two intact preintegration site amplicons shown in Figure 3 have been deposited in GenBank under accession nos. EU177095 and EU177096.

Taxonomic distribution of *M. lucifugus* transposons

To investigate the presence of related transposon families in other chiropteran or mammalian lineages, we used BLASTN or TBLASTN with nucleotide or amino acid sequence queries corresponding to each family of transposon in separate searches of custom databases representing the following taxa: Chiroptera (taxid:9397, excluding *Myotis lucifugus*), Xenarthra (taxid:9348), Afrotheria (taxid:311790), Laurasiatheria (taxid:314145), Euarch-

Table 4. Oligonucleotide primer sequences used to investigate the taxonomic distribution of the DNA transposons described

Primer	Annealing temp. (°C) ^a	Oligonucleotide sequence (5'–3')
piggyBac1_593+	58	GTCGCAGCATTGACTATAGTGATG
piggyBac1_1424-	58	CGCTTTCCTTCGCCGCAATAAATTTCC
piggyBac1_1424+	58	GGAAATTTATTGCGGCGAAGGAAAGCC
piggyBac1_2216-	58	GCACATGTAGCGTGTCTCACTCGC
piggyBac2_774+	58	CCAACGAAACTGATACACTTCCGG
piggyBac2_1536-	58	GGTGTCACTCTCCACACCATCC
piggyBac2_1536+	58	GGATGGTGTGCGAGAGTGACACC
piggyBac2_2340-	58	GCACAACTCTGCAGGCTCTCG
hAT1_1307+	58	GGTTGGCCATTATTGCTAGATATTCTGATGG
hAT1_1837-	58	GCTGACCCACTTATGATCATTGATCTGAGG
hAT1_1837+	58	CCTCAGATCAATGATCATAAGTGGGTGAGC
hAT1_2263-	58	CAATGATAATCAAATGGTACGAAACTAGAAATTAGTGAC
hAT2_309+	58	GCCCTTCTAACTCTCACTGAGCGG
hAT2_973-	59	GGTCTTTGATGTGCCTCTTGAGGTTTGG
hAT2_973+	59	CCAAACCTCAAGAGGCACATCAAAGACC
hAT2_1503-	58	GTGACCTTCAGAGATGTGCAGGAGC
hAT3_1739+	58	AATCTCAGCCGTCACCTTTGC
hAT3_2490-	58	CAAGCTTTACGACTGGATTAACAA
hAT3_2519+	58	TTCATTGAGCAAAGGGACT
hAT3_2959-	59	CATGCAGTCTCCAGTGAT
Tc1_483+	59	ATGGGATATGGAAGCCGAGGCG
Tc1_639-	59	CCAAATTTCTCACTTGCCATCTGC
Tc1_639+	59	GCAGATGGCAGAGTGAGAATTTGG
Tc1_1151-	60	CCACCTTTTGCCTTCAACTGCG
Tc2_248+	60	GGACCAGAGCATTCTGAGCCC
Tc2_928-	60	GGAGGCAGTTTGTACCATCTGCAC
Tc2_928+	60	GTGCAGATGGTACTAACTGCCTCC
Tc2_1471-	60	GTGAGATCATCTCAGTTCCTGTC
Mar_185+	60	CGTGCCAAAAAAGAGCATTGGCGG
Mar_696+	60	GGTCAACCATCAACATCGACTGC
Mar_696-	60	GCAGTCGATGTTGATGGTTGACC
Mar_1145-	60	CAGGTAATTTGTGGATACCGTCCC

Numbers in the primer name indicate their location in the respective consensus sequence. ^aAnnealing temperature for each primer pair indicated in Figure 3.

Table 5. Origin of tissue and DNA samples

Taxon	Family	Museum	Voucher ID
<i>Myotis austroriparius</i>	Vespertilionidae	Louisiana State University	M8134
<i>Myotis horsfieldii</i>	Vespertilionidae	Louisiana State University	M4424
<i>Corynorhinus rafinesquii</i>	Vespertilionidae	Louisiana State University	M8121
<i>Eptesicus furius</i>	Vespertilionidae	Texas Tech University	TK18815
<i>Kerivoula papillosa</i>	Vespertilionidae	Texas Tech University	TK152023
<i>Nycticeius humeralis</i>	Vespertilionidae	Louisiana State University	M8142
<i>Miniopterus sp.</i>	Miniopteridae	Texas Tech University	TK152087
<i>Artibeus jamaicensis</i>	Phyllostomidae	Texas Tech University	TK27682
<i>Hipposideros cervinus</i>	Hipposideridae	Texas Tech University	TK152133
<i>Rhinolophus borneoensis</i>	Rhinolophidae	Texas Tech University	TK152256
<i>Natalus stramineus</i>	Natalidae	Texas Tech University	TK101001
<i>Thyroptera tricolor</i>	Thyropteridae	Texas Tech University	TK18818
<i>Pteronotus parnellii</i>	Mormoopidae	Texas Tech University	TK21800
<i>Macroglossus sobrinus</i>	Pteropodidae	Texas Tech University	TK152047
<i>Balionycteris sp.</i>	Pteropodidae	Texas Tech University	TK152117

ontogires (taxid:314146), Mammalia (taxid:40674). We also queried the trace data from the sequencing efforts for the flying fox *Pteropus vampyrus* (7,782,409 sequences averaging 970 bp).

Oligonucleotide primers complementary to portions of the coding region of each autonomous element (Table 4) were used to survey a taxonomically diverse panel of bat genomic DNAs (Table 5) via PCR. Reaction conditions were as described previously (Ray et al. 2007) and annealing temperatures for each primer pair are found in Table 4.

Phylogenetic analyses

Amino acid sequences for known transposases from each superfamily were obtained from database searches and aligned using MUSCLE (Edgar 2004). Phylogenetic analyses were conducted using MEGA 4 (Kumar et al. 2004). Neighbor-joining trees were constructed using the equal input model with 5000 bootstrap replicates.

Acknowledgments

This work was supported by the Eberly College of Arts and Sciences at West Virginia University (D.A.R.), by start-up funds from UT Arlington (E.J.P. and C.F.) and grant R01GM77582-01 from the National Institute of Health (C.F.). N.L.C. is an investigator of the Howard Hughes Medical Institute. We thank D. Hedges for modifying his original Perl script for the CpG analyses, S. DiFazio for contributing Perl scripts to aid in processing RepeatMasker output, and R. Baker (Texas Tech University) and R. Stevens (Louisiana State University) for contributing tissue and DNA samples for this study. We thank the Broad Institute Genome Sequencing Platform and Whole genome Assembly group, K. Lindblad-Toh, and F. diPalma for making the data available.

References

- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-m., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**: 1301–1310.
- Arensburger, P., Kim, Y.J., Orsetti, J., Aluvihare, C., O'Brochta, D.A., and Atkinson, P.W. 2005. An active transposable element, Herves, from the African malaria mosquito *Anopheles gambiae*. *Genetics* **169**: 697–708.
- Avancini, R.M., Walden, K.K., and Robertson, H.M. 1996. The genomes of most animals have multiple members of the Tc1 family of transposable elements. *Genetica* **98**: 131–140.
- Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**: 251–269.

- Blair, J.E. and Hedges, S.B. 2005. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* **22**: 2275–2284.
- Borodulina, O.R. and Kramerov, D.A. 1999. Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett.* **457**: 409–413.
- Borodulina, O.R. and Kramerov, D.A. 2005. PCR-based approach to SINE isolation: Simple and complex SINEs. *Gene* **349**: 197–205.
- Calisher, C.H., Childs, J.E., Field, H.E., Holmes, K.V., and Schountz, T. 2006. Bats: Important reservoir hosts of emerging viruses. *Clin. Microbiol. Rev.* **19**: 531–545.
- Coulondre, C., Miller, J.H., Farabaugh, P.J., and Gilbert, W. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* **274**: 775–780.
- Demattei, M.V., Auge-Gouillou, C., Pollet, N., Hamelin, M.H., Meunier-Rotival, M., and Bigot, Y. 2000. Features of the mammal mar1 transposons in the human, sheep, cow, and mouse genomes and implications for their evolution. *Mamm. Genome* **11**: 1111–1116.
- Edgar, R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.
- Eick, G.N., Jacobs, D.S., and Matthee, C.A. 2005. A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (chiroptera). *Mol. Biol. Evol.* **22**: 1869–1886.
- Feschotte, C. and Pritham, E.J. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331–368.
- Feschotte, C., Zhang, X., and Wessler, S.R. 2002. Miniature inverted-repeat transposable elements and their relationship to established DNA transposons. In *Mobile DNA II* (eds. N.L. Craig et al.), pp. 1147–1158. ASM Press, Washington, DC.
- Fraser, M.J., Smith, G.E., and Summers, M.D. 1983. Acquisition of host cell DNA sequences by baculoviruses: Relationship between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedrosis viruses. *J. Virol.* **47**: 287–300.
- Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493–521.
- Gu, X. 1998. Early metazoan divergence was about 830 million years ago. *J. Mol. Evol.* **47**: 369–371.
- Hartl, D.L., Lohe, A.R., and Lozovskaya, E.R. 1997. Modern thoughts on an ancient marinere: Function, evolution, regulation. *Annu. Rev. Genet.* **31**: 337–358.
- Jehle, J.A., Nickel, A., Vlak, J.M., and Backhaus, H. 1998. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. *J. Mol. Evol.* **46**: 215–224.
- Koga, A., Iida, A., Hori, H., Shimada, A., and Shima, A. 2006. Vertebrate DNA transposon as a natural mutator: The medaka fish Tol2 element contributes to genetic variation without recognizable traces. *Mol. Biol. Evol.* **23**: 1414–1419.
- Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**: 150–163.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Leaver, M.J. 2001. A family of Tc1-like transposons from the genomes of fishes and frogs: Evidence for horizontal transmission. *Gene* **271**: 203–214.
- Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas 3rd, E.J., Zody, M.C., et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.
- Miller, D.W. and Miller, L.K. 1982. A virus mutant with an insertion of a copia-like transposable element. *Nature* **299**: 562–564.
- Miller-Butterworth, C.M., Murphy, W.J., O'Brien, S.J., Jacobs, D.S., Springer, M.S., and Teeling, E.C. 2007. A family matter: Conclusive resolution of the taxonomic position of the long-fingered bats, miniopterus. *Mol. Biol. Evol.* **24**: 1553–1561.
- Pace II, J.K. and Feschotte, C. 2007. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* **17**: 422–432.

- Piskurek, O. and Okada, N. 2007. Poxviruses as possible vectors for horizontal transfer of retrotransposons from reptiles to mammals. *Proc. Natl. Acad. Sci.* **104**: 12046–12051.
- Posada, D. and Crandall, K.A. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Pritham, E.J. and Feschotte, C. 2007. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc. Natl. Acad. Sci.* **17**: 422–432.
- Pritham, E.J., Feschotte, C., and Wessler, S.R. 2005. Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. *Mol. Biol. Evol.* **22**: 1751–1763.
- Ray, D.A., Pagan, H.J.T., Thompson, M.L., and Stevens, R.D. 2007. Bats with hATs: Evidence for recent DNA transposon activity in genus *myotis*. *Mol. Biol. Evol.* **24**: 632–639.
- Razin, A. and Riggs, A.D. 1980. DNA methylation and gene function. *Science* **210**: 604–610.
- Robertson, H.M. 1996. Members of the pogo superfamily of DNA-mediated transposons in the human genome. *Mol. Gen. Genet.* **252**: 761–766.
- Robertson, H.M. 2002. Evolution of DNA transposons in eukaryotes. In *Mobile DNA II* (eds. N.L. Craig et al.), pp. 1093–1110. ASM Press, Washington, DC.
- Ruvolo, V., Hill, J.E., and Levitt, A. 1992. The Tc2 transposon of *Caenorhabditis elegans* has the structure of a self-regulated element. *DNA Cell Biol.* **11**: 111–122.
- SanMiguel, P. and Bennetzen, J.L. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot. (Lond.)* **82**: 37–44.
- Sarkar, A., Sim, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., Robertson, H.M., and Collins, F.H. 2003. Molecular evolutionary analysis of the widespread piggyBac transposon family and related “domesticated” sequences. *Mol. Genet. Genomics* **270**: 173–180.
- Schmid, C.W. 1998. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* **26**: 4541–4550.
- Sinzelle, L., Pollet, N., Bigot, Y., and Mazabraud, A. 2005. Characterization of multiple lineages of Tc1-like elements within the genome of the amphibian *Xenopus tropicalis*. *Gene* **349**: 187–196.
- Sinzelle, L., Chesneau, A., Bigot, Y., Mazabraud, A., and Pollet, N. 2006. The mariner transposons belonging to the irritans subfamily were maintained in chordate genomes by vertical transmission. *J. Mol. Evol.* **62**: 53–65.
- Smit, A.F. and Riggs, A.D. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci.* **93**: 1443–1448.
- Stadelmann, B., Lin, L.K., Kunz, T.H., and Ruedi, M. 2007. Molecular phylogeny of New World *Myotis* (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. *Mol. Phylogenet. Evol.* **43**: 32–48.
- Swofford, D.L. 2002. PAUP*: Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, MA.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Vigdal, T.J., Kaufman, C.D., Izsvak, Z., Voytas, D.F., and Ivics, Z. 2002. Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. *J. Mol. Biol.* **323**: 441–452.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**: 973–982.
- Xing, J., Hedges, D.J., Han, K., Wang, H., Cordaux, R., and Batzer, M.A. 2004. Alu element mutation spectra: Molecular clocks and the effect of DNA methylation. *J. Mol. Biol.* **344**: 675–682.

Received September 27, 2007; accepted in revised form March 10, 2008.



Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*

David A. Ray, Cedric Feschotte, Heidi J.T. Pagan, et al.

Genome Res. 2008 18: 717-728 originally published online March 13, 2008

Access the most recent version at doi:[10.1101/gr.071886.107](https://doi.org/10.1101/gr.071886.107)

Supplemental Material <http://genome.cshlp.org/content/suppl/2008/04/08/gr.071886.107.DC1>

References This article cites 46 articles, 8 of which can be accessed free at:
<http://genome.cshlp.org/content/18/5/717.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>