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


*Genome Res.* 2008 18: 717-728; originally published online Mar 13, 2008; Access the most recent version at doi:10.1101/gr.071886.107

**Supplementary data**

“Supplemental Research Data”

http://www.genome.org/cgi/content/full/gr.071886.107/DC1

**References**

This article cites 46 articles, 20 of which can be accessed free at:

http://www.genome.org/cgi/content/full/18/5/717#References

**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:

http://www.genome.org/subscriptions/

© 2008 Cold Spring Harbor Laboratory Press
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 human 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...
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 (Repbase Update, volume 12, issue 10), a hAT transposase family member, 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 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 clad encompassing several mammalian hAT transposases (Fig. 1A). With its 3904-bp consensus, it is the longest of all of 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- 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 chiropteras 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 chiropteras, 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), opposum (Tc2\_MD2), and tenrec (Tc2\_Et), or with the human transposase-derived proteins POKG 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, Mmar1, that displays all the features of a canonical D34D mariner. The Mmar1 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 Mmar1 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
Waves of DNA transposon activity in bats

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 3'-CAGCTG-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). 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 genome 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 transposon 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...
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 Peromyscus polionotus (Family Mammidae). All three hAT families are restricted to the genus Myotis. piggyBac1_ML shares a similar distribution to the hAT elements, but positive results were obtained from the lone representatives of Minioptera, a family that was recently elevated based on 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 Laurasia-Theria, TBLASTN and BLASTN queries of chiropteran taxa with likely predates the split of Chiroptera or perhaps even Laurasia-transposons have a limited taxonomic distribution. A sister 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.

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

<table>
<thead>
<tr>
<th>Element family</th>
<th>N°</th>
<th>Mean distance</th>
<th>Substitution model</th>
<th>Std. error</th>
<th>Range</th>
<th>Age estimatea (Myr)</th>
<th>CPG Sites Included</th>
<th>CPG Sites Removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>npiyggy1_156</td>
<td>181</td>
<td>0.003</td>
<td>K80</td>
<td>0.0004</td>
<td>0.0000–0.0193</td>
<td>1.1–1.4</td>
<td>0.0021</td>
<td>K80</td>
</tr>
<tr>
<td>npiyggy1_239</td>
<td>153</td>
<td>0.0055</td>
<td>K81uf</td>
<td>0.0004</td>
<td>0.0000–0.0311</td>
<td>2.2–2.5</td>
<td>0.0029</td>
<td>K81uf+G</td>
</tr>
<tr>
<td>nhaT3_200</td>
<td>200</td>
<td>0.0155</td>
<td>Hky+G</td>
<td>0.0007</td>
<td>0.0000–0.0579</td>
<td>6.3–6.9</td>
<td>0.0079</td>
<td>Hky+G</td>
</tr>
<tr>
<td>piggyBac2</td>
<td>5</td>
<td>0.0159</td>
<td>Hky</td>
<td>0.0013</td>
<td>0.0121–0.197</td>
<td>62.7–73</td>
<td>0.0100</td>
<td>Hky</td>
</tr>
<tr>
<td>hAT3_ML</td>
<td>13</td>
<td>0.0177</td>
<td>Hky+G</td>
<td>0.0013</td>
<td>0.0374–0.233</td>
<td>69.8–90</td>
<td>0.0084</td>
<td>Hky+G</td>
</tr>
<tr>
<td>nhaT2_730</td>
<td>460</td>
<td>0.0207</td>
<td>TVM+G</td>
<td>0.0003</td>
<td>0.0733–0.484</td>
<td>86.8–9</td>
<td>0.0142</td>
<td>TVM+G</td>
</tr>
<tr>
<td>hAT1_ML</td>
<td>42</td>
<td>0.0209</td>
<td>GTR+G</td>
<td>0.0008</td>
<td>0.0132–0.393</td>
<td>8.5–9.2</td>
<td>0.0149</td>
<td>GTR+G</td>
</tr>
<tr>
<td>npiyggy2_345</td>
<td>92</td>
<td>0.0211</td>
<td>TVM+G</td>
<td>0.0016</td>
<td>0.0000–0.0599</td>
<td>8.5–9.3</td>
<td>0.0125</td>
<td>K81uf</td>
</tr>
<tr>
<td>nhaT2_1124</td>
<td>47</td>
<td>0.0235</td>
<td>TVM+G</td>
<td>0.0015</td>
<td>0.0111–0.171</td>
<td>9.3–10.6</td>
<td>0.0161</td>
<td>TVM+G</td>
</tr>
<tr>
<td>nhaT2_525</td>
<td>138</td>
<td>0.0246</td>
<td>TVM+G</td>
<td>0.0000</td>
<td>0.0069–0.520</td>
<td>10.0–10.8</td>
<td>0.0164</td>
<td>K81uf+G</td>
</tr>
<tr>
<td>nhaT2_186</td>
<td>32</td>
<td>0.0296</td>
<td>K81uf+G</td>
<td>0.0025</td>
<td>0.0096–0.707</td>
<td>11.5–13.6</td>
<td>0.0242</td>
<td>K81uf+G</td>
</tr>
<tr>
<td>nhaT2_239a</td>
<td>45</td>
<td>0.0304</td>
<td>K81uf+G</td>
<td>0.0025</td>
<td>0.0086–0.706</td>
<td>11.8–13.9</td>
<td>0.0215</td>
<td>K81uf+G</td>
</tr>
<tr>
<td>nTC1_452</td>
<td>36</td>
<td>0.0346</td>
<td>Hky+G</td>
<td>0.0019</td>
<td>0.0017–0.474</td>
<td>13.8–15.4</td>
<td>0.0201</td>
<td>Hky</td>
</tr>
<tr>
<td>nTC1_950</td>
<td>41</td>
<td>0.0365</td>
<td>Hky+G</td>
<td>0.0015</td>
<td>0.0218–0.593</td>
<td>14.8–16.1</td>
<td>0.0225</td>
<td>Hky</td>
</tr>
<tr>
<td>nMar_1285</td>
<td>69</td>
<td>0.0449</td>
<td>TVM+G</td>
<td>0.0001</td>
<td>0.0296–0.695</td>
<td>18.6–19.4</td>
<td>0.0283</td>
<td>TVM+G</td>
</tr>
<tr>
<td>nMar_1265</td>
<td>37</td>
<td>0.0473</td>
<td>TVM+G</td>
<td>0.0017</td>
<td>0.0310–0.716</td>
<td>19.3–20.7</td>
<td>0.0300</td>
<td>TVM+G</td>
</tr>
<tr>
<td>Tc2_ML</td>
<td>6</td>
<td>0.0495</td>
<td>TrN+G</td>
<td>0.0017</td>
<td>0.0452–0.543</td>
<td>20.2–21.6</td>
<td>0.0397</td>
<td>TrN</td>
</tr>
<tr>
<td>Mar_ML</td>
<td>113</td>
<td>0.0566</td>
<td>GTR+G</td>
<td>0.0011</td>
<td>0.0330–1.046</td>
<td>23.5–24.4</td>
<td>0.0324</td>
<td>GTR+G</td>
</tr>
<tr>
<td>nTC2_555</td>
<td>53</td>
<td>0.0694</td>
<td>TVM+G</td>
<td>0.0015</td>
<td>0.0464–0.993</td>
<td>28.7–30.0</td>
<td>0.0425</td>
<td>TVM+G</td>
</tr>
<tr>
<td>nTC1_527</td>
<td>25</td>
<td>0.0773</td>
<td>GTR+G</td>
<td>0.0026</td>
<td>0.00500–1.005</td>
<td>31.6–33.8</td>
<td>0.0422</td>
<td>TIM</td>
</tr>
</tbody>
</table>

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 (from Pritham and Feschotte 2007).
Figure 3. Divergence estimates and age ranges for each autonomous and nonautonomous transposon family described.
Table 2. CpG vs. non-CpG mutation ratios for each group of transposons described herein

<table>
<thead>
<tr>
<th>Element</th>
<th>CpG:nonCpG mutation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>piggyBac1_ML</td>
<td>4.48</td>
</tr>
<tr>
<td>npiggy1_156</td>
<td>2.57</td>
</tr>
<tr>
<td>npiggy1_239</td>
<td>1.56</td>
</tr>
<tr>
<td>piggyBac2_ML</td>
<td>9.40</td>
</tr>
<tr>
<td>piggy2_345</td>
<td>7.76</td>
</tr>
<tr>
<td>Myotis_hAT1</td>
<td>10.38</td>
</tr>
<tr>
<td>nhAT1_186</td>
<td>4.91</td>
</tr>
<tr>
<td>nhAT1_239a</td>
<td>4.55</td>
</tr>
<tr>
<td>hAT2_ML</td>
<td>8.83</td>
</tr>
<tr>
<td>nhAT2_730</td>
<td>10.21</td>
</tr>
<tr>
<td>nhAT2_1124</td>
<td>11.42</td>
</tr>
<tr>
<td>nhAT2_525</td>
<td>8.53</td>
</tr>
<tr>
<td>hAT3_ML</td>
<td>8.46</td>
</tr>
<tr>
<td>nhAT3_200</td>
<td>10.33</td>
</tr>
<tr>
<td>nTc1_452</td>
<td>10.06</td>
</tr>
<tr>
<td>nTc1_950</td>
<td>9.06</td>
</tr>
<tr>
<td>Mar_ML</td>
<td>8.16</td>
</tr>
<tr>
<td>nMar_1265</td>
<td>8.33</td>
</tr>
<tr>
<td>nMar_1285</td>
<td>8.24</td>
</tr>
<tr>
<td>Tc2_ML</td>
<td>8.46</td>
</tr>
<tr>
<td>nTc2_527</td>
<td>7.74</td>
</tr>
<tr>
<td>nTc2_555</td>
<td>7.92</td>
</tr>
<tr>
<td>Mean</td>
<td>7.79</td>
</tr>
</tbody>
</table>

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

...continued...
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 *HellBats* 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.

---

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

<table>
<thead>
<tr>
<th>Element family</th>
<th>Total bp analyzed</th>
<th>Total bp verified nested insertions</th>
<th>Percent nested insertions</th>
<th>Insertion type</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>piggyBac1_ML</em></td>
<td>620,634</td>
<td>214</td>
<td>0.03%</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td><em>hAT3_ML</em></td>
<td>2,649,287</td>
<td>5,029</td>
<td>0.19%</td>
<td>Yes</td>
<td>24</td>
</tr>
<tr>
<td><em>piggyBac2_ML</em></td>
<td>1,495,500</td>
<td>6,289</td>
<td>0.46%</td>
<td>Yes</td>
<td>32</td>
</tr>
<tr>
<td><em>Myotis_hAT1</em></td>
<td>4,835,096</td>
<td>27,258</td>
<td>0.56%</td>
<td><em>piggyBac1_ML</em></td>
<td>2</td>
</tr>
<tr>
<td><em>hAT2_ML</em></td>
<td>1,217,239</td>
<td>10,814</td>
<td>0.89%</td>
<td>Yes</td>
<td>126</td>
</tr>
<tr>
<td><em>Tc1_1_ML</em></td>
<td>1,750,122</td>
<td>9,619</td>
<td>0.55%</td>
<td>Yes</td>
<td>52</td>
</tr>
<tr>
<td><em>M. mariner1</em></td>
<td>1,613,024</td>
<td>25,975</td>
<td>1.61%</td>
<td><em>Myotis_hAT1</em></td>
<td>36</td>
</tr>
<tr>
<td><em>hAT3_ML</em></td>
<td>1,613,024</td>
<td>25,975</td>
<td>1.61%</td>
<td><em>hAT3_ML</em></td>
<td>3</td>
</tr>
<tr>
<td><em>piggyBac1_ML</em></td>
<td>1,613,024</td>
<td>25,975</td>
<td>1.61%</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td><em>piggyBac2_ML</em></td>
<td>1,613,024</td>
<td>25,975</td>
<td>1.61%</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td><em>hAT2_ML</em></td>
<td>1,217,239</td>
<td>10,814</td>
<td>0.89%</td>
<td>Yes</td>
<td>2</td>
</tr>
</tbody>
</table>

(Ves) VesSINE (Borodulina and Kramerov 1999).

**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.

---
Waves of DNA transposon activity in bats

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 nbaT elements when the consensus for Myotis_JaATI is used as the part of the database. Indeed, masking the genome using the Myotis_JaATI sequence produces only ∼21,000 hits greater than 100 bp. Creating a database that contains the full-length consensus and the nbaT 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 vespertilionid 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. AAPE00000000; 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 << 10−8), 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/orf/) 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).

JaAT3_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 file 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 MODELLTEST 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 duncanii, Eptesicus hobokotus, Cistugo leucuri, and Cistugo sebraii) and miønpterid (Miniopterus natalensis, M. inflatus, M. fraterculus, M. australis, and M. macroenome) 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 MODELLTEST v3.7 (Posada and Crandall 1998), TIM+G (Γ = 1.2597). We then estimated the average genetic distance between families Miøopteridae 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) into the method, we calculated a mutation rate estimate 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. australis

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 (piggyy_156) in M. australis, 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. australis. 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 amplifications 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), Euarach
ontogogtles (taxid:314146), Mammalia (taxid:40674). We also que-
ried the trace data from the sequencing efforts for the flying fox Pteropus vampyrus (7,782,409 sequences averaging 970 bp).

Oligonucleotide primers complementory 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 super-
family were obtained from database searches and aligned using MUSCLE (Edgar 2004). Phylogenetic analyses were conducted us-
ing 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 Sci-
ences 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

References


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.


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.


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