

Functional characterization of *piggyBat* from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon

Rupak Mitra^{a,b,1,2}, Xianghong Li^{a,b,1}, Aurélie Kapusta^c, David Mayhew^d, Robi D. Mitra^d, Cédric Feschotte^c, and Nancy L. Craig^{a,b,3}

^aHoward Hughes Medical Institute and ^bDepartment of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD 21205; ^cDepartment of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112; and ^dDepartment of Genetics and Center for Genome Sciences and Systems Biology, Washington University School of Medicine in St. Louis, St. Louis, MO 63108

Contributed by Nancy L. Craig, November 21, 2012 (sent for review September 4, 2012)

A revelation of the genomic age has been the contributions of the mobile DNA segments called transposable elements to chromosome structure, function, and evolution in virtually all organisms. Substantial fractions of vertebrate genomes derive from transposable elements, being dominated by retroelements that move via RNA intermediates. Although many of these elements have been inactivated by mutation, several active retroelements remain. Vertebrate genomes also contain substantial quantities and a high diversity of cut-and-paste DNA transposons, but no active representative of this class has been identified in mammals. Here we show that a cut-and-paste element called *piggyBat*, which has recently invaded the genome of the little brown bat (*Myotis lucifugus*) and is a member of the *piggyBac* superfamily, is active in its native form in transposition assays in bat and human cultured cells, as well as in the yeast *Saccharomyces cerevisiae*. Our study suggests that some DNA transposons are still actively shaping some mammalian genomes and reveals an unprecedented opportunity to study the mechanism, regulation, and genomic impact of cut-and-paste transposition in a natural mammalian host.

genome evolution | mobile genetic element

Two major classes of transposable elements that move between nonhomologous positions are known: retroelements that move via an RNA intermediate that is converted to DNA by reverse transcription and DNA-only cut-and-paste elements that move by excision of the DNA segment from a donor site, followed by integration into a target site (1). DNA-only elements are widespread in nature, being found in a wide range of prokaryotes and eukaryotes. DNA transposons or their remnants have been identified in all vertebrate genomes examined, where they are represented by diverse superfamilies and account for a substantial fraction of the genomic space, e.g., 3% in human (2), 9.1% in *Anolis* (3), and up to 25% in the African clawed frog, *Xenopus tropicalis* (4). Despite the apparent evolutionary success of DNA elements in vertebrates, we know very little about their transposition mechanism or genomic impact in these taxa. This gap in knowledge may be attributed in part to the fact that only a single, low-copy number family of vertebrate DNA elements, *Tol2*, has been demonstrated to be transpositionally active in its natural host, the medaka fish (5). There remains no direct evidence of active DNA transposons in mammals or any other amniotes. In fact, until recently, it was widely believed that all mammalian DNA transposons had gone extinct for at least 37 million years (My) (2, 6–8). However, this picture started to change several years ago when multiple waves of recent DNA transposon activity were identified through bioinformatic analyses of the genome of the little brown bat *Myotis lucifugus* (6–8). The most recently active bat transposons include members of the *hAT* and *piggyBac* superfamilies with signs of mobilization in the past few million years. However, it remained unknown whether any of these elements are actually still capable of transposition.

Here we functionally characterize the transposition activity of *piggyBat*, a member of the *piggyBac* superfamily, which likely represents the most recently active family of DNA transposons in the *M. lucifugus* genome. The founding member of the superfamily, hereafter referred to as insect *piggyBac*, was originally identified in the cabbage looper moth (*Trichoplusia ni*) (9) and has been thoroughly studied both in vivo (10) and in vitro (11). Insect *piggyBac* is known to transpose by a canonical cut-and-paste mechanism promoted by an element-encoded transposase with a catalytic site resembling the RNase H fold shared by many recombinases (11, 12). The insect *piggyBac* transposon system has been shown to be highly active in a wide range of animals, including *Drosophila* and mice, where it has been developed as a powerful tool for gene tagging and genome engineering (10). Other transposons affiliated to the *piggyBac* superfamily are common in arthropods (13) and vertebrates (14, 15) including humans (2), but none in vertebrates have been functionally examined. *piggyBac* elements present in the human genome have ceased transpositional activity about 40 My ago (16). Here we show directly that *piggyBat* is capable of transposition in bat, human, and yeast cells, thereby providing direct experimental evidence for a naturally active mammalian cut-and-paste DNA transposon. Thus, the *piggyBat* family offers an unprecedented opportunity to investigate directly the mechanism, regulation, and genomic impact of endogenous DNA transposition in a mammal.

Results

Multiple Recent *piggyBac* Families in the Bat. Previous bioinformatics mining of the *M. lucifugus* genome revealed several families of *piggyBac* elements as being most recently mobilized and the best candidates for still being active. In particular, Ray et al. (7) identified two potentially active copies (with intact ORF and identifiable target site duplication) of a family designated *piggyBat1_M1* in the 2× genome assembly available at the time. In addition, numerous short nonautonomous copies from two subfamilies called *npiggy_156* and *npiggy_239* were found to be strictly identical to each other, implying very recent transposition events and a possibly still active source of *piggyBac* transposase in the genome. We took advantage of the recent release of a much

Author contributions: R.M., X.L., A.K., D.M., R.D.M., C.F., and N.L.C. designed research; R.M., X.L., A.K., D.M., R.D.M., and C.F. performed research; R.M., X.L., A.K., D.M., R.D.M., C.F., and N.L.C. analyzed data; and R.M., X.L., A.K., D.M., R.D.M., C.F., and N.L.C. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹R.M. and X.L. contributed equally to this work.

²Present address: Hindustan Unilever Research Center, Bangalore 560066, India.

³To whom correspondence should be addressed. E-mail: ncraig@jhmi.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1217548110/-DCSupplemental.

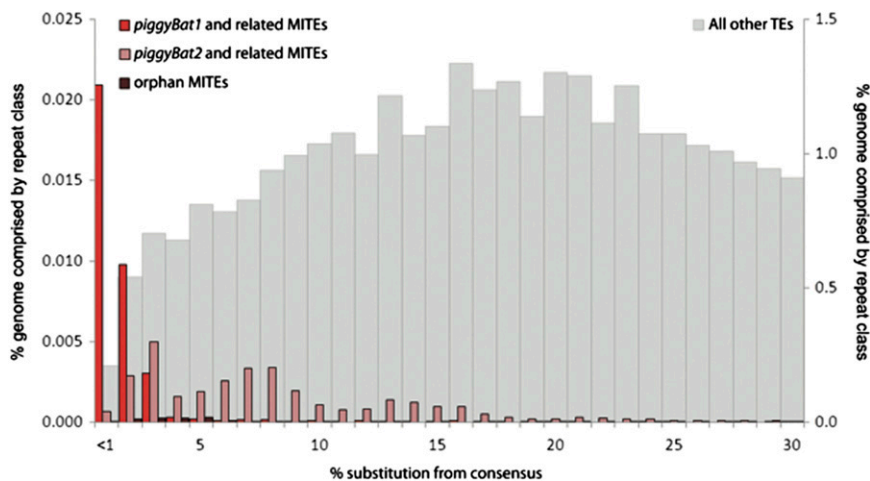


Fig. 1. *piggyBac*-like element demography versus other TEs in the *M. lucifugus* genome assembly. Percentage of divergence from consensus sequences are from RepeatMasker (47) (Methods) and corrected as in ref. 53, i.e., corrected %divergence = $-300/4 \times \ln(1 - \%divergence \times 4/300)$. MITEs, nonautonomous elements. "All other TEs" corresponds to all sequences masked in the genome as transposable elements minus all *piggyBac*-like elements.

improved genome assembly (7 \times) for *M. lucifugus* to further study the DNA transposon population of the genome and, in particular, characterize in more detail the structure and evolution of *piggyBac*-like elements. Our analysis (Methods), which involved remasking of the genome with previously identified transposons, as well as de novo mining and classification of repeats, revealed three distinct families of *piggyBac*-like elements. Each family includes transposase-encoding copies and two to seven subfamilies of nonautonomous copies or miniature inverted-repeat transposable elements (MITEs) of homogenous length and diagnostic internal sequences. Two of these families correspond to the transposase-coding families designated as *piggyBac1_Ml* and *piggyBac2_Ml* by Ray et al. (7), whereas the third one has not been previously reported. We derived consensus sequences for all families and subfamilies and used RepeatMasker to detect all copies closely related to these elements in the genome assembly and compute the percentage divergence of each copy to its consensus sequence. The results of this analysis offer a demographic profile characteristic for each family that captures the relative age of their amplification in the genome (Fig. 1). The three *piggyBac*-like families are younger than most other DNA transposon families coexisting in the bat genome, with *piggyBac1_Ml* and its related MITEs emerging as the youngest family, in agreement with previous results (7). For simplicity, we refer thereafter to *piggyBac1_Ml* as the *piggyBat* family.

Further analysis of the *piggyBat* family in the 7 \times genome assembly revealed 34 full-length copies with >99% nucleotide (nt) identity to each other, flanked by perfect 5'-TTAA-3' target site duplications (TSDs), which is required for efficient transposition of *piggyBac* (11), and containing one long ORF predicted to encode a 572-amino acid (aa) transposase. These 34 copies differ from each other at only the 3–22 nt positions (average of 12) out of 2,628 nt, their predicted transposases differ at 0–11 aa positions (average of 4), and none of these changes affect the core catalytic DDD triad (Dataset S1). Thus, we consider these 34 *piggyBat* copies as strong candidates for being autonomous elements that produce the active source of transposase. In addition, we identified a total of 1,253 MITEs with a high level (average of 97.4%) of sequence identity to each other and with 5' and 3' termini identical to full-length *piggyBat* elements, and nearly half (582) of these MITEs are still flanked by perfect TTAA TSD. Moreover, applying the neutral rate of nucleotide substitution determined by Pace et al. (11), we estimated the timing of amplification of the 34 candidate *piggyBat* autonomous elements and of the most abundant MITE family (npiggy_156) to be 0.9 and 4.5 My, respectively (Methods). Together these data suggest that a large pool of elements, including hundreds of MITEs, have been recently mobilized and might still be mobilizable by *piggyBat*

transposase. Below we validate this hypothesis functionally using a synthetic transposon system corresponding to the 1,716-bp transposase ORF (Table S1) and to the 153-bp left and 208-bp right ends of a representative full-length *piggyBat* copy identified in the genome sequence (Methods). The sequence of *piggyBat* is compared with that of insect *piggyBac* in Table S2.

***piggyBat* Can Transpose in Human and Bat Cells in Culture.** As previously described for insect *piggyBac* (17), we analyzed *piggyBat* transposition in cultured HeLa cells using a two-plasmid cotransfection assay in which a donor plasmid carries a transposon containing an antibiotic resistance marker and a helper plasmid expressing the transposase, measuring the transposase-dependent chromosomal integration of the transposon antibiotic marker. We find that the frequency of *piggyBat* integration was about 60% that of insect *piggyBac* (Fig. 2A).

We used ligation-mediated PCR to recover 98,816 transposon end-genomic junctions (Methods and Table S3) and performed high-throughput sequencing of *piggyBat* insertions in the human genome and compared the pattern of insertions to that characterized previously for insect *piggyBac* (17). We found that 92.8% of the bat *piggyBat* sites had a flanking TTAA, which is slightly lower than the 96% observed with insect *piggyBac* in the human colorectal cancer line HCT116 cells (18) and the 98% observed with insect *piggyBac* in HeLa cells (17). The distribution

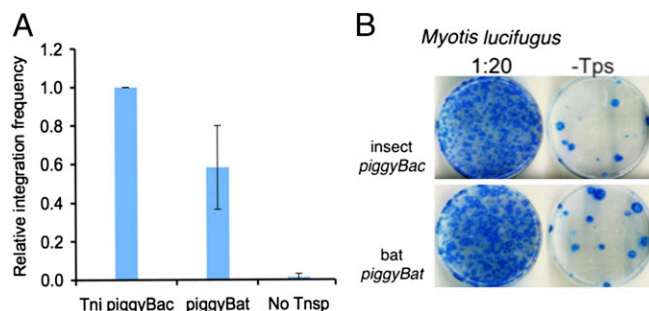


Fig. 2. Insect *piggyBac* vs. bat *piggyBat* transposition in HeLa and *M. lucifugus* cells in culture. Cells as indicated were cotransfected with a transposon donor plasmid containing either a *piggyBac* or *piggyBat* element containing a Blasticidin resistance gene and another plasmid with the cognate transposase. After 2 d, cells were treated with Blasticidin and selection continued for 18 d. Following dilution as indicated, living cells were stained with methylene blue. (A) Transposition in HeLa cells. The relative average frequencies of element transposition in five independent experiments. (B) Transposition in *M. lucifugus* fibroblast cells.

of *piggyBat* and insect *piggyBac* insertions with respect to particular human chromosomal features was also similar ($P = 0.11$, χ^2 test; Fig. 3A and B). Both transposons demonstrated biases to insert into regions of open chromatin the genome ($r^2 = 0.88$, $P = 2.27E^{-5}$; Fig. 3C) as well as regions containing a higher density of TTAA sequences ($r^2 = 0.74$, $P = 0.02$; Fig. 3D).

Using the same two-plasmid transfection system described above, we find that both bat *piggyBat* and insect *piggyBac* are also active in bat fibroblasts in culture (Fig. 2B). High-throughput capture and mapping of 4,264 de novo insertions of *piggyBat* in the bat genome show that, as in human cells, *piggyBat* insertions are enriched in the upstream region of predicted genes. The percentage of de novo *piggyBat* insertions within 5 kb of the predicted translation start sites of *Myotis lyCIFugus* genes was 11.7%, whereas that genomic space accounts for only 7.2% of TTAA targets in the genome. Interestingly, the percentage of *piggyBat* insertions in that compartment is almost identical (11.1%) to that of 2,365 preexisting *piggyBat* elements residing in the bat genome. Thus, *piggyBat* seems to favor integration of near genes, and natural selection seemingly has had a minimal impact at removing insertions from these locations.

***piggyBat* Can Excise in Yeast.** We have previously shown that insect *piggyBac* can transpose in *Saccharomyces cerevisiae* (11), and here we show that *piggyBat* can also undergo excision and integration in yeast (see *SI Text* for yeast methods). To assay excision, we used a two-plasmid system in which one plasmid expresses the transposase and the other contains a donor transposon nested within a modified version of the yeast *URA3* gene (Fig. 4A). The *URA3* gene was modified by introduction of the yeast actin intron, forming a *URA3::actin intron* gene (19). The actin intron can be efficiently spliced from mRNA of the *URA3::actin intron* gene so that a strain carrying the *URA3::actin intron* is a uracil prototroph. However, if a large DNA segment, in these experiments a mini-*piggyBat* transposon, is introduced into the actin intron, the resulting mRNA from the *URA3::actin intron::mini-piggyBat* gene is too large to be spliced, making the strain an uracil auxotroph. Thus, transposon excision can thus be measured by assaying reversion of uracil auxotrophy to prototrophy. When transposase expression is induced, we observe *piggyBat* excision at 4.1×10^{-3} Ura⁺ cells/total cells, well above the background Ura⁺ frequency of $<1 \times 10^{-7}$ Ura⁺ cells/total cells (Table 1). The frequency of *piggyBat* excision in yeast is about 10-fold lower than that previously measured for the insect *piggyBac* (11).

***piggyBat* Can Integrate in Yeast.** To analyze *piggyBat* insertion frequency in the yeast genome, we isolated chromosomal integrations of a *piggyBat* element “launched” from a plasmid. We used a two-plasmid system in which the helper plasmid expresses the transposase and the donor plasmid contains a *piggyBat*-G418^R transposon and a backbone *URA3* marker (Fig. 4B). We measured the chromosomal acquisition of *piggyBat*-G418^R in plasmid-free cells isolated by treatment with 5-fluoroorotic acid (5-FOA), which is toxic to *URA3* cells (20). The observed frequency of *piggyBat* integration using this two-plasmid system is about 3.9×10^{-4} G418^R 5-FOA^R cells/5-FOA^R cells, well above the background frequency of $<1 \times 10^{-7}$ G418^R 5-FOA^R cells/5-FOA^R cells (Table 1).

Thus, in these bulk plasmid assays, the frequency of *piggyBat* integration (3.9×10^{-4}) is about 10-fold lower than the observed excision frequency (4.1×10^{-3}), but it should be noted that the donor plasmids are different. A more direct comparison (below) shows that the frequency of *piggyBat* reintegration is actually about 40%.

***piggyBat* Frequently Reintegrates After Excision.** To more directly compare excision and reintegration, we asked what fraction of elements excised from a chromosomal *URA3::actin intron::piggyBat*-

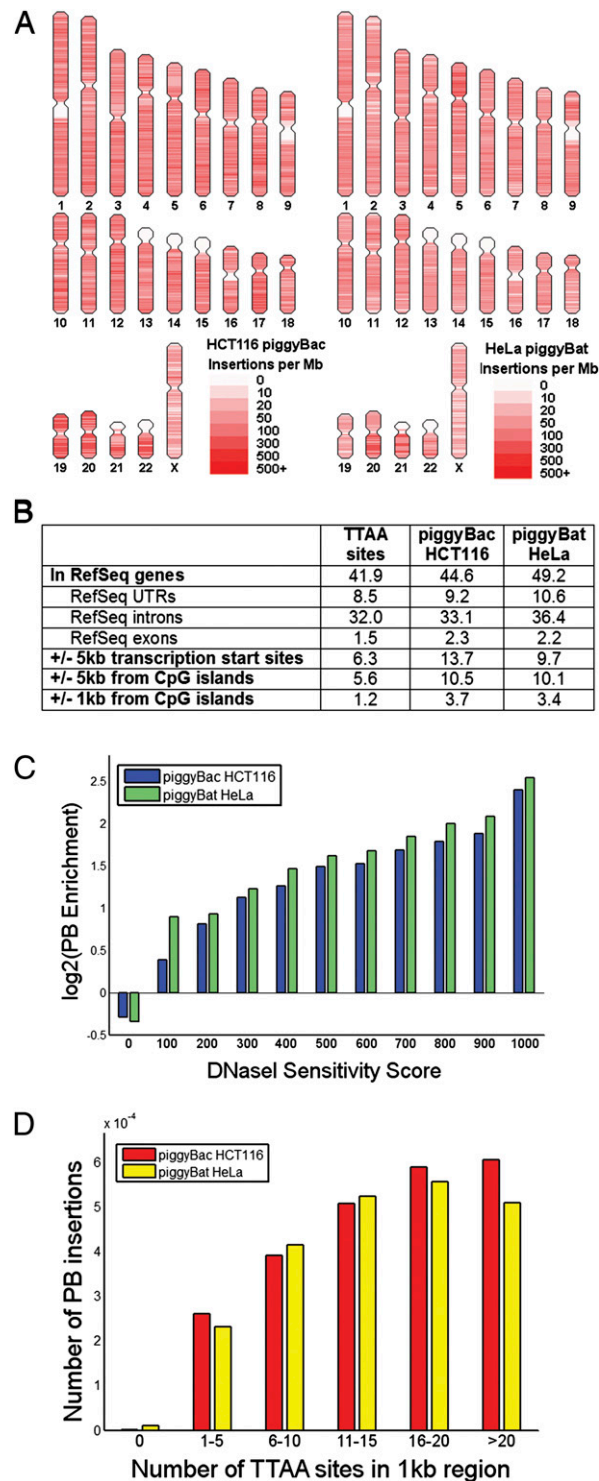


Fig. 3. Comparison of the target site selectivity of insect *piggyBac* and bat *piggyBat* in humans. (A) Patterns of the 190,000 mapped insect *piggyBac* insertions in HCT116 cells (18) and 98,800 bat *piggyBat* insertions in HeLa cells show a nonuniform distribution of insertions across the genome. (B) Regions with a higher local density of TTAA sites showed an increased preference for insect *piggyBac* insertions in HCT116 cells (18) and bat *piggyBat* insertions in HeLa cells to insert into that region. (C) Insect *piggyBac* insertions in HCT116 cells (18) and bat *piggyBat* insertions in HeLa cells are enriched in regions of open chromatin as assayed by DNase I sensitivity in terms of enrichment of insertions over the number of TTAA sites available for insertion in those regions. (D) Distributions of insect *piggyBac* insertions in HCT116 cells (18) and bat *piggyBat* insertions in HeLa cells in defined genomic regions.

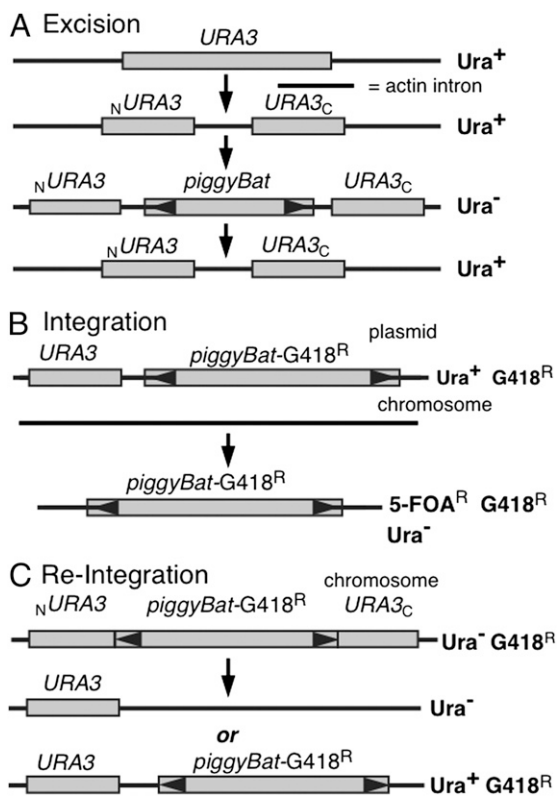


Fig. 4. *piggyBat* transposition in yeast. (A) Excision. *piggyBat* excision from a *URA3* allele was assayed by measuring reversion of uracil auxotrophy to prototrophy. (B) Integration. In the parental strain, *piggyBat* is present on a plasmid that also carries *URA3*. Integration was assayed by measuring the number of cells that retained the G418^R marker in the *piggyBat* element when the parental plasmid is excluded by treatment of the cells with 5-FOA, which is toxic to Ura⁺ cells. (C) Reintegration. In the parental strain, a *piggyBat* element carrying G418^R disrupts *URA3* such that the cells are Ura⁻. Excision is selected for by isolating cells that revert to Ura⁺, and reintegration is followed by measuring cells that continue to be G418^R.

G418^R donor site reintegrate into the yeast genome by measuring what fraction of Ura⁺ cells retained the *piggyBat*-G418 resistance marker (Fig. 4C). In one method, we compared the bulk frequencies of Ura⁺ excisants (6×10^{-4} Ura⁺ cells/total cells) vs. Ura⁺ excisants that retained the transposon-encoded G418^R marker (2.5×10^{-4} Ura⁺ G418^R cells/total cells), yielding a reintegration frequency of about 42% (Table 1). In another test, we directly assayed Ura⁺ colonies isolated from a chromosomal *URA3::actin intron::piggyBat-G418^R* donor site as to whether they were also G418-resistant, finding that of 108 Ura⁺ colonies, 40 were also G418-resistant, reflecting a reintegration frequency of 37% (Table 1), in good agreement with the 42% observed by comparison of the excision and integration frequencies.

***piggyBat* Excision Is Precise.** We sequenced the *URA3* gene in 50 cases where an *URA3::actin intron::piggyBat-G418^R* was used as a donor site, finding that in 48 of 50 cases, precise excision occurred, as has been previously shown to be true with insect *piggyBac* (21), restoring the donor site (actin intron XhoI-TTAA-*piggyBat*-TTAA-XhoI) to its preinsertion (actin intron XhoI-TTAA-XhoI) sequence. In 2 of 50 cases, the donor site was restored to the sequence of the actin intron (XhoI), likely by gene conversion using the *ACT1* gene as a template (22).

***piggyBat* Insertion Results in TTAA Target Site Duplications.** Insect *piggyBac* insertion results in TTAA target site duplication (23). To examine the integration sites used by *piggyBat* in more detail,

we isolated 20 *piggyBat* integrations into the yeast *CAN1* gene, which encodes an arginine permease, by selection for resistance to the toxic arginine analog canavanine, in a chromosomal *URA3::actin intron::piggyBat-G418^R* donor strain and sequenced both the new left and right transposon end-genomic DNA junctions. Eighteen insertions occurred at three of the seven TTAA sites in *CAN1* and were flanked by TTAA target site duplications (Table S4). The distribution of insertions at the TTAA sites was unequal, with one site containing 13 of the 18 independent insertions. The reason for this integration preference is not known.

Two *CAN1* insertions occurred at non-TTAA sites in *CAN1*: one at a CTAA target site and the other at an ATAT target site. We observed a comparable degree of target specificity for insect *piggyBac* in an analysis of about 13,500 de novo insertions in the human genome (17). One of the non-TTAA insertions in yeast, CTAA, is the same as the most common non-TTAA insertions observed for the insect *piggyBac*.

Discussion

Sequence analysis of transposable elements populating the human, mouse, rat, and dog genomes has led to the common belief that mammalian genomes lack recently active DNA transposons and that DNA transposon activity has ceased in mammalian genomes for at least the last 40 My (2, 24). Unexpectedly, a different picture emerged from the analysis of the initial genome assembly of the bat *M. lucifugus*, suggesting that several waves of DNA transposon activity occurred in the last ~40 My and persisted <1 My ago (6, 7, 25). Our bioinformatic analysis of the recently released 7x genome assembly of *M. lucifugus* corroborates these earlier investigations and confirms that the most recent waves of DNA transposition involved multiple families of *piggyBac*-like elements (7). We found that the most recent of these families, *piggyBat*, includes multiple autonomous copies (at least 34) that are >99% identical to each other, and our functional assays showed that these elements encode all of the components (transposase and *cis*-sequences) necessary and sufficient to promote transposition in bat and human cultured cells, as well as in yeast. Thus, all of the data point to *piggyBat* as a naturally active DNA transposon in a mammal.

Table 1. *piggyBat* transposition in yeast

Transposase	Frequency
<i>piggyBat</i> excision*	
Transposase	
+	4.1×10^{-3} Ura ⁺ cells/total cells
-	$<1 \times 10^{-7}$ Ura ⁺ cells/total cells
<i>piggyBat</i> integration [†]	
Transposase	
+	3.9×10^{-4} G418 ^R 5-FOA ^R cells/5-FOA ^R cells
-	$<1 \times 10^{-7}$ G418 ^R 5-FOA ^R cells/5-FOA ^R cells
<i>piggyBat</i> reintegration [‡]	
Ura ⁺ excisants	6×10^{-4} Ura ⁺ cells/total cells
Ura ⁺ G418 ^R excisants	2.5×10^{-4} Ura ⁺ G418 ^R cells/total cells
	Reintegration frequency = 42%
<i>piggyBat</i> reintegration [¶]	
Ura ⁺ excisants	108 colonies
Ura ⁺ G418 ^R excisants	40 colonies
	Reintegration frequency = 37%

*As measured by reversion of Ura⁻ cells containing *ura3::actin intron::piggyBat* to Ura⁺ cells containing *URA3::actin intron*.

[†]As measured by chromosomal acquisition of *piggyBat-G418^R* following exclusion of a donor *URA3* plasmid with 5-fluoroorotic acid.

[‡]As measured by maintenance of *piggyBat-G418^R* following selection of transposon excision from *ura3::piggyBat-G418^R*.

[¶]As measured by analysis of individual Ura⁺ colonies selected from *ura3::piggyBat-G418^R* cells for the continued presence of *piggyBat-G418^R*.

This discovery opens an unprecedented opportunity to study the mechanism, regulation, and genomic consequences of a DNA transposon family caught in the midst of invading a mammalian host. All mammalian lineages thus far examined, including primates, have been subject to DNA transposon invasions in the distant past (2, 16, 26). Some of the long-term consequences of these invasions can still be appreciated in the form of a small subset of exceptionally conserved elements coopted for cellular functions as new coding genes (27–29) or *cis*-regulatory elements (30, 31). However, the short-term impact of DNA transposon amplification remains poorly understood in mammals in particular and in eukaryotes in general. However, this class of elements is widespread in eukaryotes (28, 32), and it numerically constitutes the most frequent class of mobile elements in several well-studied organisms, such as rice (33), nematodes (34), and mosquitoes (35, 36). Recent studies in rice have revealed how a sudden burst of DNA transposition may affect host gene expression on a broad scale (37). However, such bursts are extremely difficult to catch in real time. Most genomes, and especially those of mammals, are littered with the remnants of mobile element fossils deposited dozens or hundreds of millions of years ago but contain only a tiny fraction of recently transposed elements and even fewer still capable of transposition (2, 24). Our study now reveals without ambiguity that the sequenced genome of *M. lucifugus* harbors multiple gene copies (likely in excess of 30) capable of producing active *piggyBac* transposase. Each of these transposase sources poses a formidable threat to genomic integrity because it has the potential to provoke the mobilization of several hundreds of related *piggyBac* elements currently scattered throughout the genome.

These data beg the question whether the *piggyBac* system is subject to some form of regulation and at what frequency it actually produces new insertion events in both somatic and germline tissues. Future studies shall focus on quantifying the level of *piggyBac* transposase expression in different bat tissues and examining whether host defense pathways known to repress retroelements in some mammals (e.g., small RNAs, DNA methylation) are operating to control *piggyBac* transposition. The recent development of high-throughput technologies to display and capture recent and de novo transposition events, both somatic and germline, in organisms (38, 39) should be readily applicable to estimate the frequency of *piggyBac* transposition in natural populations and to assess the amount of structural genomic variation mediated by recent and likely ongoing *piggyBac* activity.

In general, very little is known on a genome-wide scale about the consequences of transposon invasions as the elements spread or shortly after they reach fixation. Paradoxically, this is the time point where transposable elements must have the most dramatic impact on the structure and function of the genome and therefore on the evolutionary trajectory of their host, including possibly the formation of species. In this respect, the discovery of recent and likely ongoing waves of DNA transposon activity in *M. lucifugus* is intriguing because the genus *Myotis* has experienced one of the most rapid and prolific species radiation among mammals, producing >100 known extant species in the last 10–15 My (40). In particular, there is evidence that *M. lucifugus* and its closest relatives, a North American species complex called the Western long-eared *Myotis*, have experienced a very recent radiation over the last 1–1.5 My, characterized by biogeographical fragmentation and rapid morphological divergence (40–42). With the present indication of recent and likely ongoing bursts of *piggyBac* activity, this situation offers an exciting system to test the long-standing hypothesis that mobile element activity may contribute to the formation of new species (43–46).

Methods

Bioinformatics. *M. lucifugus* 7× assembly (myoLuc2) was masked for transposable elements using the RepeatMasker software (47) version 3.3.0 and RepeatMasker libraries rm-20110920 (released April 26, 2011), with the cross-match

search engine. In addition, we used a custom library containing *M. lucifugus* repeats mined de novo using Repeatscout (48) and classified with Repclass (34). Further analysis of the RepeatMasker output allowed the description of potentially autonomous and/or mobilizable *piggyBac* copies, as well as related MITEs. The number of nucleotide and amino acid differences between *piggyBac* elements were calculated with MEGA 5 (49) based on alignments generated with MUSCLE (50). Timing of *piggyBac* amplification was estimated using the neutral rate of nucleotide substitution (2.6920×10^{-9} /bp/yr) determined by Pace et al. (11), applied to the percentage of nucleotide divergence from the consensus (corrected with Jukes-Cantor distance equation). For consistency reasons (sequences names and coordinates), an additional masking of sequences from Ensembl (ftp://ftp.ensembl.org/pub/release-68/fasta/myotis_lucifugus/dna/) with all *piggyBac* elements identified in the bat genome allowed the analysis of *piggyBac1* element positions relative to predicted gene positions (ftp://ftp.ensembl.org/pub/release-68/fasta/myotis_lucifugus/cdna/) and <http://uswest.ensembl.org/biomart/martview/e86d2dbf17ed6f4e988fd22201b5d69c>).

The *piggyBac* insertions into HeLa cells and bat fibroblasts were aligned using the Novoalign software (Novocraft) with options `-r None -e 4 -l 0` to the hg19 and myoLuc2 assemblies, respectively. Only reads that could uniquely be aligned were used for subsequent analyses. The chromatin status of genomic regions was obtained from DNase I sensitivity data (51).

DNAs for Mammalian Cell Integration Assays. As described in Li et al. (17), we used a two-plasmid transposition assay, cotransfecting a donor plasmid containing a transposon carrying a GFP gene and blasticidin resistance (BsdR) cassette and a helper plasmid expressing the transposase under a cytomegalovirus (CMV) promoter to measure transposition.

The insect *piggyBac* donor plasmid contained GFP and BsdR cassettes driven by a CMV promoter, flanked by 673-bp left and 400-bp right end sequences in a pCMV/Zeo (Life Technologies) Apal, NarI fragment backbone. The insect *piggyBac* helper plasmid contained the *piggyBac* ORF cloned into the EcoRI/NotI site of pCDNA3.1 myc His A-His (Invitrogen). These DNAs have been previously described (17).

For the *piggyBac* donor plasmid, the GFP-Bsd cassette from the insect *piggyBac* mammalian donor pCMV/miniPB-GFP-Bsd was PCR amplified using primers containing SpeI sites from both ends of the GFP-Bsd cassette: 5'-GATCACTAGTGTCCGTTACATAACTTACGG and 5'-GATCACTAGTAATTCAGACATGATAAGATACATTGATGAG (SpeI sites in italics). The fragment was digested with SpeI and cloned into the plasmid-*piggyBac* TIR DNA2.0 SpeI site between the 153-bp left and 208-bp right end sequences.

In the *piggyBac* mammalian helper plasmid, the transposase was C-terminally tagged with a MYC-His tag. The *piggyBac* transposase ORF was PCR amplified from plasmid-*piggyBac* ORF DNA2.0 with a primer from the 5' end of the gene containing a HindIII site 5'-GATCAAGCTTACCATGGCGCAACACTCAGATTACTCCGACG (HindIII site is in italics and start codon is underlined) and a primer from the 3' end of the gene containing a XbaI site 5'-GATCTCTAGAATAGTTCTTCCAGCGTATGG (the XbaI site is in italics and the last codon of the transposase ORF is underlined). The PCR product was digested with HindIII and XbaI and cloned into the HindIII/XbaI sites of pCDNA3.1/myc-HisA (Invitrogen).

Growth of HeLa and Bat Cells in Culture. HeLa cells were grown in DMEM + 10% FBS + penicillin-streptomycin. The bat cells were a generous gift from Mario Cappechi (University of Utah, Salt Lake City, UT) and were a primary cell line derived from a gestation *M. lucifugus* embryo. *Myotis* cells were cultured in DMEM + 10% FBS + MEM nonessential amino acids, sodium pyruvate and penicillin-streptomycin.

Mammalian Cell Integration Assay. Integration assays in HeLa and *Myotis* cells were performed following the same procedures. Cells (2×10^5) were transfected with donor (294 nM) and helper (42 nM) plasmids with FuGENE-HD (Roche) in OPTI-MEM media (Life Technologies) according to the manufacturer's protocol. Cells transfected with donor plasmid and empty pCDNA3.1/myc-His A were the no-transposase control. After 46 h of transfection, cells were trypsinized and serially diluted in the appropriate DMEM as described above + blasticidin (3.5 μ g/mL for HeLa cells and 3.0 μ g/mL for *Myotis* cells). Fresh media with antibiotics were administered every 24 h and continued for 21 d. After 21 d, cells were fixed with 4% paraformaldehyde and stained with 0.2% methylene blue, and blue colonies were counted.

HeLa and *Myotis* Integration Libraries for High-Throughput Sequencing. Following selection for element integration with the transposon-encoded blasticidine marker, surviving HeLa and *Myotis* cells were harvested, and genomic DNA was prepared using the DNeasy Blood and Tissue Kit

(Invitrogen). Integration sites were recovered as described using ligation-mediated PCR (52). Two micrograms of genomic DNA was digested overnight with *ApoI* or *BstYI* and ligated to *ApoI* or *BstYI* linkers overnight at 16 °C. Nested PCR was then carried out under stringent conditions using end-specific primers complementary to transposon sequences and linker-specific primers complementary to the DNA linker. Primers used in this study are listed in [Table S3](#). The PCR products were AMPure XP beads purified and sequenced using the HiSeq2000 sequencing and MiSeq sequencing platforms.

- Craig NL, Craigie R, Gellert M, Lambowitz A (2002) *Mobile DNA II* (ASM Press, Washington, DC).
- Lander ES, et al.; International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921.
- Alföldi J, et al. (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477(7366):587–591.
- Hellsten U, et al. (2010) The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328(5978):633–636.
- Koga A, Suzuki M, Inagaki H, Bessho Y, Hori H (1996) Transposable element in fish. *Nature* 383(6595):30.
- Ray DA, Pagan HJ, Thompson ML, Stevens RD (2007) Bats with hATs: Evidence for recent DNA transposon activity in genus *Myotis*. *Mol Biol Evol* 24(3):632–639.
- Ray DA, et al. (2008) Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res* 18(5):717–728.
- Pace JK, 2nd, Gilbert C, Clark MS, Feschotte C (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci USA* 105(44):17023–17028.
- Fraser MJ, Smith GE, Summers MD (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(2):287–300.
- Di Matteo M, et al. (2012) PiggyBac toolbox. *Methods Mol Biol* 859:241–254.
- Mitra R, Fain-Thornton J, Craig NL (2008) piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J* 27(7):1097–1109.
- Hickman AB, Chandler M, Dyda F (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit Rev Biochem Mol Biol* 45(1):50–69.
- Bonasio R, et al. (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329(5995):1068–1071.
- Hikosaka A, Kobayashi T, Saito Y, Kawahara A (2007) Evolution of the *Xenopus* piggyBac transposon family TxpB: Domesticated and untamed strategies of transposon subfamilies. *Mol Biol Evol* 24(12):2648–2656.
- Sarkar A, et al. (2003) Molecular evolutionary analysis of the widespread piggyBac transposon family and related “domesticated” sequences. *Mol Genet Genomics* 270(2):173–180.
- Pace JK, 2nd, Feschotte C (2007) The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res* 17(4):422–432.
- Li X, et al. (2012) A resurrected mammalian hat transposable element and a closely related insect element are highly active in human cell culture. *Proc Natl Acad Sci USA* 109(1):22–27.
- Wang H, Mayhew D, Chen X, Johnston M, Mitra RD (2012) “Calling cards” for DNA-binding proteins in mammalian cells. *Genetics* 190(3):941–949.
- Yu X, Gabriel A (1999) Patching broken chromosomes with extranuclear cellular DNA. *Mol Cell* 4(5):873–881.
- Boeke JD, Trueheart J, Natsoulis G, Fink GR (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol* 154:164–175.
- Fraser MJ, Ciszczon T, Elick T, Bauser C (1996) Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol Biol* 5(2):141–151.
- Pâques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63(2):349–404.
- Cary LC, et al. (1989) Transposon mutagenesis of baculoviruses: Analysis of *Trichoplusia ni* transposon IFP2 insertions within the *FP*-locus of nuclear polyhedrosis viruses. *Virology* 172(1):156–169.
- Gibbs RA, et al.; Rat Genome Sequencing Project Consortium (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982):493–521.
- Pritham EJ, Feschotte C (2007) Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc Natl Acad Sci USA* 104(6):1895–1900.
- Smit AF, Riggs AD (1996) Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci USA* 93(4):1443–1448.
- Cordaux R, Udit S, Batzer MA, Feschotte C (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Natl Acad Sci USA* 103(21):8101–8106.
- Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41:331–368.
- Casola C, Hucks D, Feschotte C (2008) Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol Biol Evol* 25(1):29–41.
- Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9(5):397–405.
- Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 43(11):1154–1159.
- Yuan YW, Wessler SR (2011) The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci USA* 108(19):7884–7889.
- Jiang N, Feschotte C, Zhang X, Wessler SR (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr Opin Plant Biol* 7(2):115–119.
- Feschotte C, Keswani U, Ranganathan N, Guibotsy ML, Levine D (2009) Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. *Genome Biol Evol* 1:205–220.
- Nene V, et al. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316(5832):1718–1723.
- Arensburger P, et al. (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* 330(6000):86–88.
- Naito K, et al. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461(7267):1130–1134.
- Ray DA, Batzer MA (2011) Reading TE leaves: New approaches to the identification of transposable element insertions. *Genome Res* 21(6):813–820.
- Baillie JK, et al. (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479(7374):534–537.
- Stadelmann B, Lin LK, Kunz TH, Ruedi M (2007) Molecular phylogeny of New World *Myotis* (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. *Mol Phylogenet Evol* 43(1):32–48.
- Ruedi M, Mayer F (2001) Molecular systematics of bats of the genus *Myotis* (Vespertilionidae) suggests deterministic ecomorphological convergences. *Mol Phylogenet Evol* 21(3):436–448.
- Carstens BC, Dewey TA (2010) Species delimitation using a combined coalescent and information-theoretic approach: An example from North American *Myotis* bats. *Syst Biol* 59(4):400–414.
- Noor MA, Chang AS (2006) Evolutionary genetics: Jumping into a new species. *Curr Biol* 16(20):R890–R892.
- Martienssen RA (2010) Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interloid hybrids of *Arabidopsis*. *New Phytol* 186(1):46–53.
- Malone CD, Hannon GJ (2009) Molecular evolution of piRNA and transposon control pathways in *Drosophila*. *Cold Spring Harb Symp Quant Biol* 74:225–234.
- Maheshwari S, Barbash DA (2011) The genetics of hybrid incompatibilities. *Annu Rev Genet* 45:331–355.
- Smit A, Hubley R, Green P (1996–2010) RepeatMasker open-3.0. Available at www.RepeatMasker.Org.
- Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.
- Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
- Sabo PJ, et al. (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 3(7):511–518.
- Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD (2007) HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res* 17(8):1186–1194.
- Waterston RH, et al.; Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520–562.