Rolling-Circle Transposons Catalyze Genomic Innovation in a Mammalian Lineage

Jainy Thomas1, Caleb D. Phillips2, Robert J. Baker2, and Ellen J. Pritham1,*

1Department of Human Genetics, University of Utah
2Department of Biological Sciences and Museum, Texas Tech University
*Corresponding author: E-mail: pritham@genetics.utah.edu.

Accepted: September 8, 2014

Abstract

Rolling-circle transposons (Helitrons) are a newly discovered group of mobile DNA widespread in plant and invertebrate genomes but limited to the bat family Vespertilionidae among mammals. Little is known about the long-term impact of Helitron activity because the genomes where Helitron activity has been extensively studied are predominated by young families. Here, we report a comprehensive catalog of vetted Helitrons from the 7 × Myotis lucifugus genome assembly. To estimate the timing of transposition, we scored presence/absence across related vespertilionid genome sequences with estimated divergence times. This analysis revealed that the Helitran family has been a persistent source of genomic innovation throughout the vespertilionid diversification from approximately 30–36 Ma to as recently as approximately 1.8–6 Ma. This is the first report of persistent Helitron transposition over an extended evolutionary timeframe. These findings illustrate that the pattern of Helitron activity is akin to the vertical persistence of LINE retrotransposons in primates and other mammalian lineages. Like retrotransposition in primates, rolling-circle transposition has generated lineage-specific variation and accounts for approximately 110 Mb, approximately 6% of the genome of M. lucifugus. The Helitrons carry a heterogeneous assortment of host sequence including retroposed messenger RNAs, retrotransposons, DNA transposons, as well as introns, exons and regulatory regions (promoters, 5′-untranslated regions [UTRs], and 3′-UTRs) of which some are evolving in a pattern suggestive of purifying selection. Evidence that Helitrons have contributed putative promoters, exons, splice sites, polyadenylation sites, and microRNA-binding sites to transcripts otherwise conserved across mammals is presented, and the implication of Helitron activity to innovation in these unique mammals is discussed.

Key words: Helitron, transposable element, gene capture, gene duplication, Vespertilionidae bat family, retrogene.

Introduction

The largest fraction of multicellular eukaryotic genomes is composed of autonomously replicating DNA, called transposable elements (TEs) (for review, Pritham 2009). TEs engender variation directly through insertion and excision and indirectly through recombination (for review, Kidwell and Lisch 2002). The sequence generated through transposition and TE-mediated recombination is initially unconstrained and can serve as a source of genomic innovation (for review, Fedoroff 2012). In addition to self-replication and the generation of nonautonomous copies, TEs mediate the duplication of constrained host genes and regulatory sequences. This process can occur at the DNA level (through retroposition) (for review, Kaessmann et al. 2009) as well as at the DNA level through transduction or recombination (for review, Kaessmann 2010). Novel chimeric genes can result from the fusion of TE-encoded genes or acting sequences with host gene fragments (e.g., Cordaux et al. 2006; for review, Feschotte and Pritham 2007) as well as from the fusion of host sequences from unlinked locations (Moran et al. 1999; Lal et al. 2003; Jiang et al. 2004; Kawasaki and Nitasaka 2004; Gupta et al. 2005; Lai et al. 2005; Morgante et al. 2005; Zabala and Vodkin 2005; Hoen et al. 2006; Wang and Dooner 2006; Wang et al. 2006; Xing et al. 2006; Xu and Messing 2006; Jameson et al. 2008; Sela et al. 2008; Sweredoski et al. 2008; Xiao et al. 2008; Damert et al. 2009; Du et al. 2009; Hancks et al. 2009; Langdon et al. 2009; Rogers et al. 2009; Yang and Bennetzen 2009a; Elrouby and Bureau 2010; Han et al. 2013). Among TEs, rolling-circle elements (Helitrons) mediate gene chimera formation and have been proposed to be important players in new gene formation (e.g., Morgante 2006; Fan et al. 2008; Thomas et al. 2014; for review, Bennetzen 2005; Morgante 2006).

The genome of the little brown bat, Myotis lucifugus is home to a large diversity of both retrotransposons and DNA...
transposons (Pritham and Feschotte 2007; Ray et al. 2007, 2008; Pagan et al. 2012; Zhuo et al. 2013). Two features of the *M. lucifugus* genome (and related genomes) are unusual among mammals. The first feature is the diversity of recently active DNA transposon families (younger than 40 Myr) (Ray et al. 2008), a pattern that is distinct from most other mammals, which have few if any active DNA transposons (most families are older than 40 Myr) (Pace and Feschotte 2007). The DNA transposon families appear to have entered the genome in waves at different time points during the last approximately 40 Myr (Pritham and Feschotte 2007; Ray et al. 2007, 2008) at least in some cases by horizontal transfer (Pace et al. 2008; Thomas et al. 2010). The second feature is the presence of *Helitrons*, which is in marked contrast to the dearth of related sequences in other mammals (except for two highly degraded gene fragments of the *Helitron* encoded Rep/Helicase protein found in the genome of platypus) (Pritham and Feschotte 2007; Thomas et al. 2010, 2011).

*Helitrons* are widespread in plants and invertebrates often contributing to a sizable fraction of the genome (as much as ~6.62%) (fig. 1). They have also been described from some microbial eukaryotes, fungal, and vertebrate genomes but have not reached high copy number (Poulter et al. 2003; Thomas et al. 2008; Thomas et al. 2010, 2014). *Helitrons* from bats began getting fixed approximately 36 Ma (Pritham and Feschotte 2007); therefore, most copies are substantially older than their counterparts in other eukaryotes (it has been estimated that most of the maize *Helitrons* transposed <250,000 years ago) (Feschotte and Pritham 2009; Yang and Bennetzen 2009a). The discovery of *Helitrons* in the genome of the little brown bat (Pritham and Feschotte 2007; Thomas et al. 2011) affords the unique opportunity to understand the tempo of transposition in a genome that evolves slower (Yi et al. 2002; Bininda-Emonds 2007) than the previously studied plant and insects (Petrov and Hartl 1998; Bennetzen 2009). The vast array of mammalian genomic resources helps to measure for the first time the long-term impact of *Helitron* activity on an array of evolutionary features, such as speciation, morphology, and genome organization.

*Helitrons* display features that are atypical of other eukaryotic DNA transposons. This includes the putative rolling-circle-like transposition mechanism, unique structure, and lack of target site duplication (TSD) (Kapitonov and Jurka 2001). The putative autonomous *Helitrons* carry a single open reading frame (ORF) (Rep/Helicase) with homology to proteins involved in rolling-circle replication encoded by other forms of mobile DNA (Kapitonov and Jurka 2001). The sequence features are limited to a ‘TC’ on the 5’-end and a short stem loop structure followed by a terminal ‘CTRR’ on the 3’-end (Kapitonov and Jurka 2001). *Helitrons* display a very complex genetic structure, thus the classification is much different than for other DNA transposons (Du et al. 2009; Yang and Bennetzen 2009b). Members of the same family share at least 80% identity over the last 30 bases and belong to the same subfamily when they also share at least 80% identity over the first 30 bases. Exemplars have unique (>20% divergence) intervening sequence (Yang and Bennetzen 2009a). The heterogeneity of *Helitron* units and the minimal sequence conservation makes curation difficult to automate. Some programs (Du et al. 2008; Yang and Bennetzen 2009b) work well to identify putative *Helitrons* that display the structural features described above but verification of prior mobility is necessary for accurate curation and to avoid false positive calls. Evidence can include the presence of multiple copies with the same termini at different positions and/or the identification of empty sites (insertion free sites).

*Helitrons* capture genic sequence at the DNA level (Lai et al. 2003, 2005; Brunner et al. 2005; Gupta et al. 2005; Morgante et al. 2005; Wang and Dooner 2006; Xu and Messing 2006; Cultrone et al. 2007; Hollister and Gaut 2007; Tempel et al. 2007; Fan et al. 2008; Jameson et al. 2008; Sweredoski et al. 2008; Du et al. 2009; Langdon et al. 2009; Yang and Bennetzen 2009a; Wicker et al. 2010; Dong et al. 2011; Coates et al. 2012; Han et al. 2013; Thomas et al. 2014) through a poorly understood process, which is likely in part a result of the bypass of the 5’- or 3’-end and the low
sequence requirements for transposition. Helitrons have been dubbed “exon-shuffling machines” due to the propensity for gene capture and the resulting sequence heterogeneity (Feschotte and Wessler 2001). Gene capture has been well documented in the maize genome where Helitron activity has resulted in the transduction and shuffling of tens of thousands of gene fragments into novel genetic units (protogenes), which in some rare instances have been further amplified through Helitron transposition (Lal et al. 2003; Brunner et al. 2005; Gupta et al. 2005; Lal and Hannah 2005; Morgante et al. 2005; Wang and Dooner 2006; Xu and Messing 2006; Jameson et al. 2008; Du et al. 2009; Yang and Bennetzen 2009a). Well-supported gene fragment acquisition has also been reported in lepidopterans, fungi, and a few other plants (Cultrone et al. 2007; Hollister and Gaut 2007; Tempel et al. 2007; Fan et al. 2008; Sweredoski et al. 2008; Langdon et al. 2009; Wicker et al. 2010; Han et al. 2013). Previously we reported one example of gene capture in the M. lucifugus genome by HelibatN3, which has amplified the promoter, 5′-untranslated region (UTR), first exon, and partial first intron of the NUBPL gene to a thousand copies (Pritham and Feschotte 2007). Although there is substantial data supporting the significance of gene duplication to evolutionary innovation in bats (Seim et al. 2013; Zhang et al. 2013), the importance of Helitrons in this process remains to be investigated.

The birth of a functional gene duplicate or new gene is not likely to be immediate but rather to proceed through intermediate and reversible protegene stages (Carvunis et al. 2012). For de novo genes, there is an evolutionary continuum and no strict boundary between the different stages (Carvunis et al. 2012). Several studies in Drosophila have identified novel functional genes from protegene duplicates (Chen et al. 2007; Yang et al. 2008; Ding et al. 2010; Kogan et al. 2012) that are associated with a subtype of Helitron (Thomas et al. 2014). It is noteworthy that most of the gene captures and chimera genes are evolutionarily neutral (Juretic et al. 2005); however, occasionally a small fraction get transcribed and translated suggestive of functional potential (Hanada et al. 2009). The null hypothesis is that the Helitron amplified protegenes are evolutionarily neutral.

The previously estimated period of activity of the Helibat family was approximately 30–36 Ma (Pritham and Feschotte 2007). This was based on the analysis of a few subfamilies. However, the extent and impact of Helitron transposition and gene chimera formation across species has not been previously investigated. Studies of Helitrons in maize, Arabidopsis, rice (Hollister and Gaut 2007; Sweredoski et al. 2008; Du et al. 2009; Yang and Bennetzen 2009a), and lepidopteran genomes (Han et al. 2013) revealed much about the impact of recent Helitron transposition but are not informative about long-term dynamics due to a comparatively high rate of genomic turnover (Drosophila ~60 times faster than mammals; ~14.3 Myr [Petrov and Hartl 1998]; maize estimated <2 Myr [Bennetzen 2009]) and low average age of Helitrons (0.25–5 Myr old) (Hollister and Gaut 2007; Sweredoski et al. 2008; Du et al. 2009; Yang and Bennetzen 2009a; Han et al. 2013). The M. lucifugus and related genomes on the other hand have a slower mutation rate and longer unconstrained DNA half-life (884 Myr) (Petrov and Hartl 1998), thus are ideal to study long-term activity and impact of Helitrons.

Here, we leverage the recent release of a high-quality 7× genome assembly of M. lucifugus to carry out the first comprehensive characterization of Helitrons in a mammalian genome. We also analyzed RNA-sequencing (RNA-seq) data from an adult M. lucifugus salivary gland to develop an understanding of Helitron-mediated structural alterations of a bat transcriptome. Most profoundly, we provide the first evidence that Helitron activity spans the diversification of the Vespertilionidae family, with recent activity around approximately 1.8–6 Ma, a pattern that resembles the vertical persistence of LINE retrotransposons in primates (Hormozdiari et al. 2013). This is the first indication that Helitrons are capable of long-term persistence and can maintain activity over extended evolutionary periods, which is in sharp contrast to cut-and-paste DNA transposons in animals (Pace and Feschotte 2007; for review, Robertson 2002; Schack et al. 2010; Huang et al. 2012). In addition, we report that Helitrons have sequentially captured gene fragments, cDNAs, and regulatory regions creating lineage-specific gene chimeras. The genomic impact of this horizontally transferred TE family on the structure of the genome and transcriptome is discussed in light of the unique transposition mechanism, long-term persistence, and propensity for gene shuffling.

**Materials and Methods**

**Identification and Classification of Helitrons**

A complementary combination of de novo and structure-based approaches (Repeatscout 1.0.2 [Price et al. 2005] and Helsearch [Yang and Bennetzen 2009b]) was employed to identify Helitrons from the M. lucifugus genome (GL429767–GL433173, 3,407 scaffolds, from GenBank at National Center for Biotechnology Information [NCBI]). The repeat families identified by Repeatscout were assembled (Sequencher 4.7; >90% identity and 100-bp overlap) and Repclass (Feschotte et al. 2009) was employed to identify Helitrons from the M. lucifugus genome.

Identification of Host Genomic Sequences within Helitrons

To identify gene fragments captured by Helitrons, a representative of each family, subfamily, and exemplar was queried...
using nucleotide-based searches (BLASTn, default parameters) against all mammalian genomes deposited at NCBI. Considering the slow rate of sequence decay (Yi et al. 2002; Bininda-Emonds 2007) and the absence of Helitrons in most mammals (Pritham and Feschotte 2007; Thomas et al. 2011), the only significant hits represented regions homologous to non-Helitron regions conserved in other mammalian genomes. The resulting hits (E-value \( < 10^{-04} \), >50 bp) were then queried against the human gene information from the UCSC (University of California–Santa Cruz) genome browser to further annotate the host gene fragments as promoter, UTRs, exon, or intron. Empty sites were identified to validate the mobility of the Helitrons containing host gene fragments and to confirm the boundaries of those belonging to novel families (supplementary materials and methods, Supplementary Material online). The gene fragments within Helitron are considered of retrogene origin, if the gene fragment is characterized by multiple exons, which are devoid of introns. In some cases, poly A tail and TSDs flanking the gene were identified. Helitron containing retrogenes with a poly A tail and TSDs were considered as retroposition to the Helitron, where as retrogenes with no identifiable TSDs were considered as capture of pre-existing retrogenes. For these cases, empty sites were sought to confirm the retroposition rather than capture.

Simulation analyses were performed to identify bias in ease of identification of any particular region of a gene (upstream/promoter, 5’-UTR, coding exons, introns, and 3’-UTR) using homology-based methods. In addition, the influence of Helitron position was assessed (detailed in supplementary materials and methods, Supplementary Material online). To assess the nature of the promoter regions captured by Helitrons, the human homologs were extracted using the UCSC genome browser. The chromatin state segmentation data of the corresponding captured promoter sequences from nine human cell types were observed using UCSC genome browser (Hoffman et al. 2013).

Estimation of Copy Number and Abundance of Helitrons in the genome

To estimate Helitron copy numbers, the first and last 30 bp of all Helitrons were queried to the M. lucifugus genome using RepeatMasker (v 4.0.2) (Smit et al. 1996–2010) and hits greater than 80% identical were counted. To estimate the proportion of the genome occupied, the Helitron library was used to Repeatmask the M. lucifugus genome (supplementary materials and methods, Supplementary Material online). Copy numbers of Helitrons containing host sequences were extracted from the RepeatMasker output by counting Helitrons covering greater than 50% of the total length of the query Helitron.

Estimation of Length of Activity of Helitrons in M. lucifugus Genome

A subset of the Helitron library (containing host sequences) was blasted (BLASTn) against E. fuscus and M. davidii genomes available at NCBI to identify the gene capture (amplification) events unique to each lineage. Age of such capture events was estimated based on the presence/absence of insertions in both species and from the estimated divergence time (Miller-Butterworth et al. 2007). In some cases, age of Helitrons was also calculated based on the average divergence from the consensus and the estimated neutral mutation rate (Ray et al. 2008) (supplementary materials and methods, Supplementary Material online). To further support the continued activity of Helitrons, we sought to identify the putatively autonomous Helitron in the M. lucifugus genome. The Helibat1 consensus (Pritham and Feschotte 2007) was used to blast (BLASTn) against M. lucifugus genome, individual copies were extracted, and the presence of intact ORFs was examined using ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html, last accessed September 22, 2014).

Tests for Selection

The parental is defined as the source gene from which the retroposed copies arose. If the capture of the retrogene occurred prior to the divergence of M. lucifugus, the putative parental was reconstructed by generating a consensus from the aligned orthologous copies. Upon identification of copies with an intact ORF, analyses were performed to estimate the selection pressure on the coding regions. Synonymous (dS) and nonsynonymous (dN) substitution rates were calculated by codeml in the PAML package (Goldman and Yang 1994) by using PAL2NAL (Suyama et al. 2006) with the default parameters. Further, a maximum-likelihood ratio test was performed if the dN/dS ratio was less than 0.5 (indicative of purifying selection) to further test the statistical significance (Yang 1998; Betrán et al. 2002).

Identification of Helitron Containing Transcripts from the Submandibular Gland Transcriptome

The illumina RNA-seq reads (75 bp) from the submandibular gland of M. lucifugus (Phillips et al. 2014) (SRP031492) were assembled using Trinity (Grabherr et al. 2011) using both de novo and genome-guided approaches (supplementary materials and methods, Supplementary Material online). Transcripts with minimum expression abundance (fragments per kilobase of transcript per million mapped reads [fpkm] \( > 1 \), \( \geq 200 \) bp) were queried using the Helitron library using BLASTn (v2.2.21). Hits with E-values \( < 10^{-4} \) over at least 30 bp were manually examined and categorized based on how Helitrons altered the structure of the transcript (supplementary materials and methods, Supplementary Material online). The coding potential calculator (Kong et al. 2007) and BLAST tools were used to assess the coding capacity of the
transcripts, and noncoding transcripts were further categorized based on the criteria reviewed in Ilott and Ponting (2013) (supplementary materials and methods, Supplementary Material online).

**Results**

**Classification and Abundance of Helitrons in M. lucifugus**

A library of 645 unique Helitrons (with precisely defined termini), comprising 44 families (supplementary table S1, Supplementary Material online), 47 subfamilies, and 534 exemplars was compiled by employing multiple computational tools (Materials and Methods). To validate the boundary and mobility of the element, orthologous or paralogous empty sites were identified for each family and subfamily reported (supplementary materials and methods and fig. S1, Supplementary Material online). The average size of the elements was 804 bp (±703 SD) and the individual copies ranged from 123 to 5,503 bp in length. Copy numbers for each Helitron subfamily were estimated by counting the 5’-(first 30 bp) and 3’-ends (last 30 bp) that had at least 80% identity to the query (Materials and Methods, supplementary materials and methods, Supplementary Material online). In total, 132,717 (5’)- and 159,643 (3’)-ends were identified indicating the presence of at least 132,717 Helitrons. In total, Helitron-derived sequences accounted for 109.8 Mb of DNA (fig. 1) (5.8% of the ~1.89 Gb assembly).

**Identification of Gene Fragments Captured and Amplified by Bat Helitrons**

To identify Helitrons that have captured gene fragments, we leveraged the absence of these elements in mammals (outside of vespertilionid bats) (Pritham and Feschotte 2007; Thomas et al. 2011) and the slow evolving nature of mammalian genomes (Yi et al. 2002; Bininda-Emonds 2007) compared with other plants and insects (Petrov and Hartl 1998; Bennetzen 2009). This cross-species comparative approach allowed the demarcation of sequences homologous to those found in other mammalian genomes but embedded within Helitrons in M. lucifugus. The genomic fragments within the Helitrons were aligned to the well-annotated human genome for annotation (promoter, UTR, intron, exon, etc.). Using this approach we found that 110 of 645 unique Helitrons carry host sequences from 54 different genes (supplementary tables S2 and S3, Supplementary Material online) and from 16 distinct nongenic regions conserved in other mammalian genomes (supplementary table S4, Supplementary Material online).

Helitrons that carry host sequences (protogenes) were amplified to 12,382 copies (supplementary tables S2–S4, Supplementary Material online). Examination of the captured regions revealed a higher frequency of sequences corresponding to the 5’-end of the gene (promoters, 5’-UTR, first coding exon and first intron) compared with other genic regions (internal or last exon, other introns, 3’-UTR) (fig. 2 and supplementary table S2, Supplementary Material online). In addition, Helitrons containing promoters, 5’-UTRs, or the first exon/intron tend to amplify to higher copy number (ranging from 4,690 to 5,505 copies) than those containing other genic regions (28–1,381 copies) (fig. 2). This pattern was not accounted for by a bias in our methods of detecting potentially more conserved regulatory regions (supplementary data set S1 and results, Supplementary Material online) nor does it reflect a distribution bias for Helitron insertions in the genome (supplementary data set S2 and results, Supplementary Material online).

We further investigated the nature of the chromatin state in nine human cell types generated by the ENCODE project (Hoffman et al. 2013) for the sequences corresponding to captured promoters in Helitrons (15 distinct genes). The corresponding human sequences displayed chromatin marks consistent with promoter and frequently enhancer activity in at least one of the cell lines tested (supplementary table S5, Supplementary Material online). This analysis is in agreement with the human annotation and suggests that Helitrons carrying these sequences given the correct genomic context might be able to drive transcription. Transcription is necessary for a protogene to gain function and be selected upon.
Evidence of Prolonged Helitron Activity in the Vespertilionid Bat Lineage

By using cross-species comparative genomics, we estimate that there were at least 11,688 Helitron generated protogenes fixed prior to the divergence of Eptesicus and Myotis (fig. 3A). In addition, we found that at least 392 protogenes are *E. fuscus*-specific, 668 are *Myotis*-specific (shared by both *M. lucifugus* and *M. davidii* but absent in *E. fuscus*) (fig. 3A and supplementary table S6, Supplementary Material online), and 26 insertions are *M. lucifugus*-specific. These data confirm the activity of the Helibat family prior to (~25 Ma) (Pritham and Feschotte 2007; Thomas et al. 2011) and provide evidence for continued activity after the diversification of the *Myotis* lineage (12 Ma).

This cross-species analysis also revealed sequential capture of host gene fragments during the diversification of the Vespertilionidae. For instance, we identified three exemplars with different sets of gene fragments, each of which were specific to a different branch of the bat phylogeny (as shown in fig. 3A). The capture of the 5′-UTR, first exon, and partial first intron of a homolog of the human transmembrane BAX inhibitor motif containing 4 (*TMBIM4*) by *HelibatN217.1* occurred in the ancestor of the three bats (fig. 3A–C). After the divergence of *Eptesicus* and *Myotis*, a second capture event occurred in the ancestral *Myotis* lineage where a partial retrogene of the transforming acidic coiled-coil containing protein 3 (*TACC3*) (fig. 3D and E) was inserted upstream of the captured *TMBIM4* fragment (annotated as *HelibatN22*, 17 copies). Further analysis of copies of the Helitron containing the *TACC3* in the *M. lucifugus* genome led to the identification of a third, recently amplified subfamily *HelibatN541* (fig. 4), which carries a fragment of *TACC3* retrogene (fig. 4A). This fragment can be readily aligned to the *TACC3* retrogene capture discussed above; however, whether it is the result of an independent capture or a deletion of the previously mentioned gene capture is unclear. We identified 26 closely related copies (98–99% identity to the subfamily consensus) of *HelibatN541* in the *M. lucifugus* assembly but none in the *M. davidii* assembly (fig. 4B–D). We also estimated that *HelibatN541* transposed approximately 1.8–6 Ma based on sequence divergence (0.9% ± 0.49 SD) and neutral substitution rate (Ray et al. 2008) (see supplementary materials and methods, Supplementary Material online), which would indeed postdate the divergence of the two *Myotis* species (~12 Ma) (fig. 3A). To further support the possibility of recent activity, we examined the current *M. lucifugus* genome assembly to identify potentially intact autonomous elements containing Zn finger motifs, Rep and Helicase domains. We did not find an intact element, but we found that a 1,496 amino acids (aa) ORF encoding Zn finger motifs, Rep and Helicase domains could be reconstituted by introducing three changes in a copy (GL429822.1: 1008959–1014346).

In summary, we found that most *Helitron* exemplars arose at least 25 Ma but report several lines of evidence that support the transposition of the *Helibat* family persisted over most of the course of vespertilionid bat evolution (as recently as 1.8–6 Ma) engendering lineage-specific genetic variation.

How Do Helitrons Capture Gene Fragments?

Several models have been proposed to explain gene acquisition by Helitrons (Feschotte and Wessler 2001; Brunner et al. 2005; Kapitonov and Jurka 2007; Tempel et al. 2007; Lal et al. 2009). Feschotte and Wessler (2001) proposed that capture of host sequences occurs when there is inefficient recognition of the termination signal of a Helitron close to the captured region leading to the transduction of adjacent downstream host sequence and later termination by a random signal. Updating this model, Tempel et al. (2007) proposed that another Helitron inserted nearby acts as the termination signal, resulting in a chimeric Helitron with host gene sequences in between. The parental copies of all genes that were captured by Helitrons were examined to identify related Helitrons (with similar ends as that of the *Helitron* with gene capture) adjacent to the parental captured region (supplementary materials and methods, Supplementary Material online). In most cases, we failed to identify related Helitrons near the captured region (26 cases). This could be because a related Helitron was never nearby, or that it was there but was not fixed or it was subjected to lineage sorting so present in some lineages but not others. For 20 cases, the assembly was poor in the region examined. However, in three cases (*HelibatN539, HelibatN217.1*, and *HelibatN549*), we found nearly identical 3′-ends of Helitrons adjacent to the parental gene. In the parental copy of *HelibatN539*, we also found a nearly identical 5′-end in the same orientation flanking the captured host region (*KCNQ5* gene) (supplementary fig. S2, Supplementary Material online). Thus, it appears that in this case a composite transposition event occurred using the 5′-end of one Helitron copy and the 3′-end of the other copy resulting in the creation of a compound transposon *HelibatN539* (supplementary fig. S2 and table S2, Supplementary Material online). The result is a novel chimeric Helitron and the capture of the intervening region, which supports the end bypass model (Feschotte and Wessler 2001; Tempel et al. 2007). It should be noted that although *HelibatN539* is present in *E. fuscus* as well as *Myotis*, the progenitor *Helitron* at the 3′-end could only be identified in the *E. fuscus* genome (supplementary fig. S2, Supplementary Material online).

Helitron-Mediated Amplification of Nested Retrogens

The retroposition of messenger RNA (mRNA) to generate retrogens is a common mechanism of gene duplication in mammalian genomes (for review, Zhang 2003; Kaessmann et al. 2009). We identified four distinct instances of mRNAs that had been reverse transcribed into Helitrons and further propagated (118 copies; supplementary table S3, Supplementary
Genomic Innovation in a Mammalian Lineage

Helibats

TMBIM4

43Myr

TACC3

25Myr

(668)

(392)

7

1

3

5

9

13

M. davidii

Eptesicus fuscus

M. lucifugus

Miniopteridae

Myotis davidii

HelibatN22

Miniopteridae

Eptesicus fuscus

Genome Helitron

Fig. 3.—Continuous activity and sequential capture of gene fragments by Helitrons. (A) The phylogenetic relationship of the three vespertilionid bat genome sequences publicly available (used in this study) and the sister family Miniopteridae. Activity of Helitrons is limited to the vespertilionid bat lineage and is estimated to have begun around 30–36 Ma (red lighting bolt) (Pritham and Feschotte 2007). The timing of divergence is represented at each node (in Myr) (Miller-Butterworth et al. 2007; Stadelmann et al. 2007; Gu et al. 2008). The number within the parenthesis represents the lineage-specific Helitron copies with host sequences. The lineage-specific insertions in the M. davidii and Eptesicus genome are underestimates as we did not thoroughly analyze the Helitron content of those genomes. (B) A cartoon structure of the Helitron containing TMBIM4 (involved in apoptosis inhibition [Saravat et al. 2013] gene fragment, HelibatN217.1) (C) An alignment of the TMBIM4 fragment within HelibatN217.1 to that of the host fragment from the mouse lemur, bat (parental), and human genome. The captured region shares 87% identity over 223 bp (excluding gaps) with the captured region spanning the last four exons and the 3′-UTR of the gene shared approximately 78% sequence identity over 265-bp excluding gaps. The color of the lines above the alignment indicates the region of the gene (e.g., red line shows 5′-UTR, blue line shows the coding exons, purple line shows the 3′-UTR, and pink line denotes the intron).
Material online). These secondary insertions display the hallmark of retrogenes: They lack introns, terminate by a stretch of adenines, and are flanked by TSDs consistent with LINE-1-mediated retroposition events (Esnault et al. 2000). In all four cases, we were able to identify exemplars that lacked the retrogene (empty sites) suggesting the retroposition of those genes into the Helitrons rather than capture of the retrogene elsewhere with TSDs (supplementary fig. S1, Supplementary Material online). These secondary insertions display the hallmark of retrogenes: They lack introns, terminate by a stretch of adenines, and are flanked by TSDs consistent with LINE-1-mediated retroposition events (Esnault et al. 2000). In all four cases, we were able to identify exemplars that lacked the retrogene (empty sites) suggesting the retroposition of those genes into the Helitrons rather than capture of the retrogene elsewhere with TSDs (supplementary fig. S1, Supplementary Material online). This is the first report of such retrogene propagation by Helitrons or, to our knowledge, any DNA transposon. Although all four cases involved 5′-truncated insertions (a common outcome in LINE-1-mediated retroposition [Zingler et al. 2005]), in two of those cases only a small region of the 5′-UTR is missing (ribosomal protein, large, P0 [RPLP0] [HelibatN424] and phosphoglycerate dehydrogenase [HelibatN544]) (supplementary table S3, Supplementary Material online).

Our analysis revealed a novel mechanism by which Helitron transposition in conjunction with the LINE-1-mediated retroposition of mRNA can create proteogens by bringing together a promoter and retroposed sequences. Retrogenes typically lack promoter regions and integration near a promoter region, may facilitate transcription (for review, Zhang 2003; Kaessmann et al. 2009; Kaessmann 2010). For instance, HelibatN211 (73 copies) carries the promoter, 5′-UTR, exon, and intron of SRPK1 (fig. 5A and B). An mRNA of the RPLP0 was retroposed into the captured promoter region (339 bp upstream of 5′-UTR) of SRPK1 in the HelibatN211 forming a new exemplar, HelibatN424 (fig. 5C and D). HelibatN424 was amplified to six copies before the diversification of Eptesicus and Myotis. Whether the HelibatN424 proteogene is transcribed from this promoter remains to be explored, as the SRPK1 promoter typically drives expression in the direction away from the retrogene (Amin et al. 2011). However, the process illustrates one mechanism by which retrogenes could be brought into close proximity to prefabricated promoter regions, gaining the sequences necessary to support transcription.

Evidence of Potential Function

To identify potentially functional Helitron-amplified genes, we measure the ratio of dN/dS (Yang 1998; Betrán et al. 2002) on the coding region of copies that had complete intact ORFs compared with the parental gene (see Materials and Methods). Four of the 36 copies of the TCF3 (E2A) fusion partner (TFPT) retrogene amplified by HelibatN102 (supplementary table S2, Supplementary Material online) have intact ORFs as described. Presence of multiple orthologous copies in M. davidii indicates the capture of the retrogene occurred before the divergence of the two bat species. To take that to account, we constructed a super consensus.
Fig. 5.—RPLP0 mRNA retroposed into the exompr HelibatN211. (A) The structure of the HelibatN211 exemplar, which contains SRPK1 gene fragment. (B) Alignment of the SRPK1 (a single copy human gene known to play a regulatory role in intron splicing [for review, Giannakouros et al. 2011]) gene fragment within Helitron to human and walrus. Yellow and orange lines show the 30 bp at the 5′ gene (without introns) from Helitron CCATGCACAAATTTCATGCACTGGGCCACTAG

Helitron

Human

Walrus

1.10...20...30...40...50...60...70

*** *** *** *** *** *** *** *** *** ***

Helitron

Human

Walrus

...1410...1420...1430...1440...1450...1460...1470

...1550...1560...1570...1580...1590...1600...1610

Helitron

Human

Walrus

...2320...2330...2340...

C

TC

SRPK1

TSD

RPLP0

TSD

CTAG

3′

D

******** * * ******** * ******** * ** ******** ** * ****

Helitron

Human

Mouse

Parental

1.10...20...30...40...50...60...70

******** * * ******** * ******** * ** ******** ** * ****

Helitron

Human

Mouse

Parental

500...510...520...530...540...550...560

******** * * **** * ******** * ******** * ** ******** ** * ****

Helitron

Human

Mouse

Parental

1060...1070...1080...1090...1100...1110...1120

******** * * **** * ******** * ******** * ** ******** ** * ****

Helitron

Human

Mouse

Parental

1130...1140...1150...1160...
(from eight orthologous copies) to represent the parental gene that was retroposed. Among the four copies with intact ORFs, the dN/dS ratio for one copy is significantly less than 0.5 (0.39, P < 0.05). The orthologous Helitron amplified retrogene in the M. davidii genome has a higher dN/dS ratio (0.65) in addition to the presence of multiple premature stop codons suggesting that the observed signals of selection are not due to pre-existing constraints. To further understand the pattern of selection pressure on other copies within the M. lucifugus genome that do not have an intact ORF, a similar analysis was performed. Only two of the nearly complete 25 TPTF copies have dN/dS ratio less than 0.5 (P < 0.05). These two copies had stop codons within first 110 aa and have a methionine amino acid after the stop codon. However, the signatures of purifying selection (supplementary table S7, Supplementary Material online) suggest that the duplicates may be functional.

### Impact of Helitrons on the Transcriptome

To assess the impact of Helitrons on the evolution of transcript structure, we analyzed transcriptome data generated for the salivary gland of M. lucifugus (Materials and Methods). A total of 29,493 distinct transcripts including isoforms (≥ 200 bp) were assembled from RNA-seq data (~30 million paired end reads; with fpkm > 1). Of these transcripts, 417 (1.4%) contain a Helitron sequence. Further inspection of these transcripts revealed that Helitron insertions have altered every aspect of transcript structure, including the site of transcription initiation, the length and sequence of the 5'UTR and 3'UTR and exon boundaries (fig. 6A–E, Materials and Methods, supplementary materials and methods, Supplementary Material online). These transcripts include those that are potentially coding (38%) and noncoding (62%) (supplementary data set S3, Supplementary Material online). Helitrons are found integrated into both orthologs and paralogs of human long noncoding RNAs (lncRNAs) (fig. 6F and supplementary data set S3, Supplementary Material online) or as components of lncRNAs not previously described in other species (198 transcripts; supplementary table S8 and data set S3, Supplementary Material online). Thus, Helitron insertions have altered the structure of a fraction of both coding (mostly 3’-UTR) as well as noncoding transcripts.

Due to the preponderance of Helitrons found in the 3’-UTR of transcripts (fig. 6) and the multiple roles of the 3’-UTR in gene regulation (for review, Grzybowska et al. 2001), we sought to identify regulatory motifs carried by the subset of Helitrons found in the 3’-UTR of miRNAs. Using PITA (microRNA [miRNA] target finder tool) (Kertesz et al. 2007) and based on the human miRNAs deposited in miBase (Kozomara and Griffiths-Jones 2011), a total of 338 predicted miRNA-binding sites (8 mer seeds with no mismatches and with minimum energy [supplementary materials and methods, Supplementary Material online] [Kertesz et al. 2007])

![Fig. 6.](image)

The subset of adult salivary gland transcript containing Helitrons. (A) Categorization of transcripts that have human homologs. In XY%, X corresponds to the number of transcripts that fall in that category (alternative transcripts containing the same Helitron are counted only once). Y denotes the percent of the total transcripts (excluding the alternate transcripts) altered by a Helitron in each category. Note that if a transcript can be included in multiple categories, it is counted as a member of all applicable categories. (B) to (I) are examples of various categories. (B) A Helitron is part of the 3’-UTR. It is inserted in the intron in the 5’-UTR of the DEDD gene, interrupting the splicing of the intron. (C) Helitron introduces a premature stop codon (shown as a star). It is inserted in the intron before the 3’-UTR of YWHAQ gene and interrupts the splicing of the intron. (D) A Helitron provides a cryptic splice site and fuses with a transcript from the ZMYM4 gene. The Helitron provides an initiation codon and a few codons in the predicted ORF of the alternative transcript of ZMYM4 gene. (E) A gene fragment within Helitron provides a cryptic...
for 203 miRNAs were identified in Helitron sequences embedded in 52 transcripts (~72%, fig. 6G and supplementary data set S4, Supplementary Material online). Supporting the putative function, 209 of the predicted sites were perfectly conserved in either or both the E. fuscus and M. davidii orthologs. To determine whether these predicted miRNA-binding sites existed in Helitrons or whether they evolved after insertion into the 3'-UTR, we compared their sequences with the consensus generated from multiple Helitron copies using majority rule. These analyses revealed that 213 of the 338 predicted sites were not identified in the consensus suggesting that they evolved after insertion in the 3'-UTRs. These predicted sites are also identified at orthologous positions in E. fuscus (25) and in M. davidii (110) suggestive of their putative function. To further assess the potential biological relevance of these predicted miRNA-binding sites, the presence of bat homologs of the corresponding precursor miRNAs (pre-mir) and their predicted miRNA-binding sites, the presence of bat homologs of the corresponding precursor miRNAs (pre-mir) and their expression in the salivary gland transcriptome was assessed (supplementary materials and methods, Supplementary Table S9 and data set S3, Supplementary Material online). The presence of bat homologs of the corresponding precursor miRNAs (pre-mir) and their expression in the salivary gland transcriptome was assessed (supplementary materials and methods, Supplementary Material online). We found that 61 of 193 pre-mirs (multiple mature miRNAs can arise from a single precursor [Griffiths-Jones et al. 2006]) have homologs in the M. lucifugus genome and seven of these are expressed (>1 fpkm) in the salivary gland transcriptome (supplementary data set S4, Supplementary Material online). Together these data indicate a possible role of Helitrons in seeding of miRNA-binding sites throughout the M. lucifugus transcriptome. To identify other putative regulatory motifs, Helitron-containing transcripts were analyzed using UTRScan, which predicts the presence of functionally characterized motifs in UTRs (Grillo et al. 2010). Several Mushashi-binding elements were predicted in Helitrons residing in the 3'-UTR of 16 transcripts (20 sites) (fig. 6H and supplementary table S9, Supplementary Material online). The Mushashi protein is involved in the posttranscriptional regulation of several key developmental genes (Phillips et al. 2013; for review, MacNicol et al. 2008). We found that the bat ortholog of human Mushashi-2 (MS2H) is expressed in the submandibular gland at a relatively high level (15 fpkm); thus, it is conceivable that this protein could be involved in the regulation of those genes through binding to the Mushashi-binding elements predicted in Helitrons. In addition, we found that Helitrons provide novel polyadenylation sites for two transcripts originating from genes (OTUD3 and TOR1B) that are highly conserved across vertebrates (fig. 6I and supplementary table S9 and data set S3, Supplementary Material online). Each of these examples suggests that some of the Helitrons residing in the 3'-UTRs of M. lucifugus miRNAs may have acquired regulatory roles, although these remain to be functionally tested.

**Discussion**

The little brown bat genome affords the opportunity to look at the dynamics and consequences of Helitron amplification at a much deeper evolutionary time period relative to the plant and insect genomes previously studied (Hollister and Gaut 2007; Sweredoski et al. 2008; Du et al. 2009; Yang and Bennetzen 2009a; Han et al. 2013). Helitron activity has generated approximately 6% of the M. lucifugus genome (~110 Mb; fig. 1), which percentagewise is close to that described for the lepidopteran, Heliconius melpomene (6.6%; 17.1 Mb) and is more than approximately 2% (45 Mb) of the B73 inbred line of the maize (Yang and Bennetzen 2009a) and approximately 4.2% (20 Mb) of the silkworm genome (fig. 1) (Han et al. 2013). The total amount of bat Helitron DNA (110 Mb) is approximately 2 × more than what is seen in the other genomes. In M. lucifugus, many Helibat subfamilies are amplified to several hundreds of copies (supplementary tables S2 and S4, Supplementary Material online) where as in maize very few subfamilies have gained such high copy number (Du et al. 2009; Yang and Bennetzen 2009a). Another key difference is that the Helibat subfamilies range in age, from approximately 1.8 to approximately 36 Myr old (fig. 3). This is in striking contrast to maize where most of the insertions are thought to be less than 250,000 years old, and in silkworm where the peak of Helitron activity was within the past 2 Myr (Yang and Bennetzen 2009a; Han et al. 2013). Indeed some Helibat subfamilies and exemplars are lineage-specific, thus Helibat activity has garnered genomic structural variation throughout the Vespertilionidae family. This extended period of activity in concert with the relatively slow rate of mammalian evolution (Yi et al. 2002; Bininda-Emonds 2007) in comparison with other plants and insects contributes to the large amount of Helitron-generated DNA found in the M. lucifugus genome today.

Long-term activity is unorthodox for DNA transposons in mammalian genomes, which display a different transposition mechanism (cut-and-paste vs. rolling-circle) and are typically characterized by a burst of rapid proliferation followed by mass extinction (Pace and Feschotte 2007; for review, Robertson 2002; Schack et al. 2010; Huang et al. 2012). This pattern of long-term activity of cut-and-paste DNA transposons has been proposed based on the observation of vertical diversification in plants (Lisch et al. 2001; Feschotte et al. 2003).
a pattern not seen with animal cut-and-paste DNA transposon families. In animals, this pattern is more similar to the vertically persistent non-Long Terminal Repeat (LTR) retrotransposon proliferation well characterized in mammalian genomes (Hormozdiari et al. 2013). Perhaps coincidentally, the amount of DNA that has been generated by Helitrons in the *M. lucifugus* genome is comparable to the amount of DNA generated by L1 retrotransposition in the human lineage over the same period of evolution (Khan et al. 2006). Helitron activity has created several hundreds of insertions that are either Myotis-specific or Eptesicus-specific (figs. 3 and 4 and supplementary table S6, Supplementary Material online). This clearly demonstrates the role of Helitrons in inducing genomic structural variation. These lineage-specific insertions could be leveraged as markers to reconstruct the phylogeny of vespertilionid bats, which has been recalcitrant to classical phylogenetic approaches due to their recent and rapid diversification (Lack and Van den Bussche 2010; Larsen et al. 2012). It is possible that Helitrons could be active in some vespertilionid bats, as we were able to reconstruct a putative autonomous copy with three changes. Therefore, bat Helitrons might provide an interesting system to dissect the transposition and gene capture mechanisms in an experimental setting.

In addition to structural variation, bat Helitron activity has led to genomic innovation through capture, duplication, and fusion of host gene sequence (the generation of protogenes). A total of 12,382 Helitron elements have amplified host sequences (supplementary tables S2–S4, Supplementary Material online). We limited our study to well-vetted repeat families (confirmed with an orthologous or paralogous empty site). In addition, our detection method relies on homology with other mammalian genome sequences and gene captures were documented and annotated only when the sequence was conserved in the human genome (Materials and Methods). The Helitrons carrying host sequences account for approximately 0.6% of the *M. lucifugus* genome (supplementary tables S2–S4, Supplementary Material online). This amount of gene duplication is remarkable given that mammalian protein-coding exons typically account for approximately 1.5% of the genome (Lander et al. 2001) and suggest that these protogenes could have profound evolutionary implications. In a pattern similar to what is seen in maize or insect (Lai et al. 2003, 2005; Brunner et al. 2005; Gupta et al. 2005; Morgante et al. 2005; Jameson et al. 2008; Yang and Bennetzen 2009a; Barbaglia et al. 2012; Han et al. 2013), we report that some Helitron-generated protogenes are transcribed or under purifying selection. Furthermore, the number of Helitrons containing host sequences in the *M. lucifugus* genome (12,382) is lower than maize (at least 20,000) (Yang and Bennetzen 2009a). In maize, the number of gene fragments acquired per exemplar is reported to range from 1 to 9 (Du et al. 2009; Yang and Bennetzen 2009a), where as in *M. lucifugus*, it is 1–2. In addition, the length of Helitrons in maize varies from 202 to 35,925 bp (Yang and Bennetzen 2009a) (average length of 4,616 bp) (Sweredoski et al. 2008), where, as in *M. lucifugus* the range is 123–5,503 bp (average 804 bp ±703 SD). The high frequency of gene capture is not observed in either the rice or *Arabidopsis* genomes where the average length of each element is 441 and 950 bp, respectively (Sweredoski et al. 2008). However, it is interesting to note that rice has large number of Pack-MULEs (>3,000) which have captured and amplified gene fragments from more than 1,000 genes (Jiang et al. 2004).

One striking finding was the capture of promoter regions from 15 genes by different exemplars and subfamilies, which then amplified to high copy numbers (4,690 copies dispersed throughout the genome) (fig. 2 and supplementary table S2, Supplementary Material online). The presence of a promoter (if functional) could promote transcription of regions carried in proximity, which is the first step toward becoming a functional gene. Indeed, the Helitrons carrying these promoters often had gene fragments from other regions of the genome captured in a sequential manner at the DNA level or at the RNA level as we describe for HelibatN211 (fig. 5). Although we do not know the evolutionary implication of this example, it serves to highlight the exceptional ability of Helitrons to create novel protogene chimeras that may be capable of immediate transcription. Capture of nearly full-length genes was reported in maize (Xu and Messing 2006; Jameson et al. 2008), where a few of the amplified copies had intact ORFs (with introns) (Xu and Messing 2006; Jameson et al. 2008), which were transcribed in multiple tissues (Jameson et al. 2008). Here, we show for the first time that cDNA copies of mRNA are integrated into Helitrons and then further amplified by Helitron transposition, some of which are under purifying selection indicative of their potential function (supplementary tables S3 and S7, Supplementary Material online).

The evidence of functional constraints acting on some of the retrogenes (supplementary table S7, Supplementary Material online) amplified by bat Helitrons supports the view that protogene formation through gene capture at both the DNA as well as the RNA level is likely an important process to bat genome evolution (for review, Kaessmann et al. 2009; Kaessmann 2010). Indeed, Helitrons have promoted the amplification of three of the gene families (*TFPT*, *RPLPO*, and *NARF*) identified in analysis of the *M. brandtii* and *M. davidii* genomes (Seim et al. 2013; Zhang et al. 2013). It has been proposed that the gene duplicates have played a pivotal role in the evolution of vespertilionid bats (Zhang et al. 2013).

To begin to understand the pattern of transcription and the usage of captured regulatory regions, we analyzed an RNA-seq data set from adult salivary gland. Our finding that 1.4% of the transcripts (>1 fpkm) began with, contained or ended in a Helitron was remarkably consistent with a study of full-length silkworm cDNA (123 of 8,654; 1.4%) collected from multiple tissues at different developmental stages (Han et al. 2013).
We leveraged the extensive resources of transcription data from human to understand the ancestral state of the transcript (before Helitron insertion) and to interpret the variation engendered by the presence of the Helitron. Helitrons contributed putative novel transcription start sites, splice sites, polyadenylation signals, as well as 3'-UTR sequences with predicted miRNA-binding sites and known regulatory motifs to transcripts otherwise conserved across mammals (fig. 6; supplementary data sets S3 and S4 and table S9, Supplementary Material online). In addition, we found evidence that Helitron insertions have diversified the sequence of otherwise well-conserved lncRNAs (fig. 6), which supports the hypothesis that TEs are important contributors to the evolution of the lncRNA repertoire of mammals (Kelley and Rinn 2012; Kapusta et al. 2013) (supplementary materials and methods, table S8, and data set S3, Supplementary Material online).

Together these data underscore the potential role of Helitrons in remodeling gene expression at both transcriptional and posttranscriptional levels, adding to the growing body of evidence that mobile elements represent major players in the lineage-specific tinkering of regulatory networks (for review, Feschotte 2008; Bourque 2009; Rebollo et al. 2012). The limitations of the study include a single adult somatic tissue transcriptome, low coverage, and the lack of strand-specific RNA-seq data. For these reasons, we observe here in regards to the influence of these elements on the transcriptome is likely the tip of the iceberg.

Genomic incompatibilities are likely to be fundamental to the establishment of reproductive barriers in the diverse and species-rich Vespertilionidae family (407 species, 48 genera) (Simmons 2005; Baker and Bradley 2006). Diversification within the family is extensive with the oldest extinct species having diverged approximately 40 Ma (lines of fossil evidence [Horacek 2001; Miller-Butterworth et al. 2007]) and the most recent having diverged approximately 1 Ma (Stadelmann et al. 2007). Several genomic features that can be promoted by TE activity have been proposed to contribute to the evolution of reproductive isolation (for review, Carroll 2005; Böhne et al. 2009; Rebollo et al. 2010; Romero et al. 2012; Platt et al. 2014). Some of these intrinsic factors include chromosomal rearrangements and variation in the gene structure and expression (for review, Wolf et al. 2010; Nosil and Schluter 2011). The continued transposition of Helitrons (~1.8–36 Ma) along with the transcription and dispersal of genic fragments, including regulatory regions (promoters, UTRs), exons, exon–intron splicing sites, and retrogenes in a lineage-specific fashion could have played a role in generating incompatibilities. It is intriguing that Helitron activity mirrors the diversification of the Vespertilionidae family and is tempting to speculate that Helitron-mediated genomic incompatibilities, if not causal to speciation, might have played a role in the reinforcement of reproductive isolation.

**Supplementary Material**

Supplementary materials and methods, results, figures S1 and S2, tables S1–S9, and data sets S1–S4 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org).

**Acknowledgments**

The authors thank DIAG and XSEDE for providing the resources to perform Trinity and the Trinity support group for technical support. They also thank Cedric Feschotte, Aurelie Kapusta and Raymond Maライフon-Borja for critical reviewing of the manuscript, Lixing Yang for running the Helssearch program, and Aurelie Kapusta for providing the scripts to analyze the RepeatMasker output. This work was supported by startup funds from the University of Utah to E.J.P.

**Literature Cited**

Amin EM, et al. 2011. WT1 mutants reveal SRPK1 to be a downstream angiogenesis target by altering VEGF splicing. Cancer Cell 20: 768–780.

Baker RJ, Bradley RD. 2006. Speciation in mammals and the genetic species concept. J Mammal. 87:643–662.

Barbaglia AM, et al. 2012. Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics 190:965–975.

Bennetzen JL. 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev. 15:621–627.

Bennetzen JL. 2009. Maize genome structure and evolution. In: Bennetzen JL, Hake SC, editors. Handbook of maize: genetics and genomics. New York: Springer. p. 190–200.

Betrán E, Thornton K, Long M. 2002. Retroposed new genes out of the X chromosome. Science 297:1854–1859.

Bininda-Emonds OR. 2007. Fast genes and slow clades: comparative rates of molecular evolution in mammals. Evol Bioinform Online. 3:59–85.

Böhne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff JN. 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. Chromosome Res. 16:203–215.

Bourque G. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. Curr Opin Genet Dev. 19:607–612.

Brummer S, Pea G, Rafalski A. 2005. Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize. Plant J. 43:799–810.

Carroll SB. 2005. Evolution at two levels: on genes and form. PLoS Biol. 3:e245.

Carvunis AR, et al. 2012. Proto-genes and de novo gene birth. Nature 487:370–374.

Chen ST, Cheng HC, Barbash DA, Yang HP. 2007. Evolution of hydra, a recently evolved testis-expressed gene with nine alternative first exons in Drosophila melanogaster. PLoS Genet. 3:e107.

Coates BS, Hellmich RL, Grant DM, Abel CA. 2012. Mobilizing the genome of Lepidoptera through novel sequence gains and end creation by non-autonomous Lep1 Helitrons. DNA Res. 19:11–21.

Cordaux R, Uditz 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 U S A. 103:8101–8106.

Cultrone A, Dominguez YR, Drevet C, Scacciozch C, Fernandez-Martin R. 2007. The tightly regulated promoter of the xan4 gene of Aspergillus nidulans is included in a helitron. Mol Microbiol. 63:1577–1587.
Dong Y, et al. 2011. Structural characterization of Helitrons and their step-wise capturing of gene fragments in the maize genome. BMC Genomics 12:609.

Du C, Caronna J, He L, Dooner HK. 2008. Computational prediction and molecular confirmation of Helitron transposons in the maize genome. BMC Genomics 9:51.

Du C, Fefelova N, Caronna J, He LM, Dooner HK. 2009. The polychromatic Helitron landscape of the maize genome. Proc Natl Acad Sci U S A. 106:19916–19921.

Elrouby N, Bureau TE. 2010. Bs1, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize. Plant Physiol. 153: 1413–1424.

Ensault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Genet. 24:363–367.

Fan C, et al. 2008. The subtelomere of chromosome 3 short arm as a hot bed of new gene origination in rice. Mol Plant. 1: 839–850.

Fedoroff NV. 2012. Presidential address. Transposable elements, epigenetics, and genome evolution. Science 338:758–767.

Feschotte C. 2008. Opinion—transposable elements and the evolution of regulatory networks. Nat Rev Genet. 9:397–405.

Feschotte C, Kewani U, Ranganathan N, Gubosky 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.

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

Feschotte C, Pritham EJ. 2009. A cornucopia of Helitron-related transposable elements in maize contain portions of multiple pseudogenes. Plant Mol Biol. 57:115–127.

Han MI, et al. 2013. Identification and evolution of the silkworm Helitrons and their contribution to transcripts. DNA Res. 20:471–484.

Hanada K, et al. 2009. The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. Plant Cell 21:25–38.

Hancock DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH. 2009. Exon-trapping mediated by the human retrotransposon SVA. Genome Res. 19:1983–1991.

Hoen DR, et al. 2006. Transposon-mediated expansion and diversification of a family of ULP-like genes. Mol Biol Evol. 23:1254–1268.

Hoffman MM, et al. 2013. Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. 41:827–841.

Hollister JD, Gaut BS. 2007. Population and evolutionary dynamics of Helitron transposable elements in Arabidopsis thaliana. Mol Biol Evol. 24:2515–2524.

Horacek I. 2001. On the early history of vespertilionid bats in Europe: the Lower Miocene record from the Bohemian Massif. Lynx (Praha) 32: 123–154.

Hormozdari F, et al. 2013. Rates and patterns of great ape retrotransposition. Proc Natl Acad Sci U S A. 110:13457–13462.

Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. Annu Rev Genet. 46:651–675.

Ilott NE, Ponting CP. 2013. Predicting long non-coding RNAs using RNA sequencing. Methods 63:50–59.

Jameson N, et al. 2008. Helitron mediated amplification of cytochrome P450 monoxygenase gene in maize. Plant Mol Biol. 67:295–304.

Jiang N, Bao ZR, Zhang XY, Eddy SR, Wessler SR. 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature 431: 569–573.

Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE. 2005. The evolutionary fate of MULE-meditated duplications of host gene fragments in rice. Genome Res. 15:1292–1297.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20:1313–1326.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet. 10:19–31.

Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A. 98:8714–8719.

Kapitonov VV, Jurka J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet. 23:521–529.

Kapusta A, et al. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 9:e1003470.

Kawasaki S, Nitasaka E. 2004. Characterization of Tpn1 family in the Japanese morning glory: EnSpm-related transposable elements capturing host genes. Plant Cell Physiol. 45:933–944.

Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 13:R107.

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. Nat Genet. 39:1278–1284.

Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 16:78–87.

Kidwell MG, Lisch D. 2002. Transposable elements as sources of genomic variation. In: Craig R, Gellert M, Gellert M, Gellert M, Gellert M, eds. The origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 9:e1003470.

Kawasaki S, Nitasaka E. 2004. Characterization of Tpn1 family in the Japanese morning glory: EnSpm-related transposable elements capturing host genes. Plant Cell Physiol. 45:933–944.

Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 13:R107.

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. Nat Genet. 39:1278–1284.

Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 16:78–87.

Kidwell MG, Lisch D. 2002. Transposable elements as sources of genomic variation. In: Craig R, Gellert M, Gellert M, Gellert M, Gellert M, eds. The origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 9:e1003470.
Genomic Innovation in a Mammalian Lineage

Kong L, et al. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35:W345–W349.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39: D152–D157.

Lack JB, Van den Bussche RA. 2010. Identifying the confounding factors in resolving phylogenetic relationships in Vespertilionidae. J Mammal. 91: 1435–1448.

Lai JS, Li YB, Messing J, Dooner HK. 2005. Gene movement by Helitron transposons contributes to the haplotype variability of maize. Proc Natl Acad Sci U S A. 102:9068–9073.

Lal S, Oetjens M, Hannah LC. 2009. Helitrons: enigmatic abductors and mobilizers of host genome sequences. Plant Sci. 176:181–186.

Lal SK, Giroux MJ, Brendel V, Vallejos CE, Hannah LC. 2009. The maize genome contains a Helitron insertion. Plant Cell 15:381–391.

Langdon T, et al. 2009. Fragments of the key flowering gene GIGANTEA are associated with helitron-type sequences in the Pooidaeae grass Lolium perenne. BMC Plant Biol. 9:70.

Larsen RJ, et al. 2012. Genetic diversity of Neotropical Myotis (Chiroptera: Vespertilionidae) with an emphasis on South American species. PLoS One 7:e46578.

Lisch DR, Freeling M, Langham RJ, Choy MY. 2001. Mutator transposase is widespread in the grasses. Plant Physiol. 125:1293–1303.

MacNicol AM, Wilczynska A, MacNicol MC. 2008. Function and regulation of the mammalian Musashi mRNA translational regulator. Biochem Soc Trans. 36:528–530.

Miller-Butterworth CM, et al. 2007. A family matter: conclusive resolution of the taxonomic position of the long-fingered bats, Miniopterus. Mol Biol Evol. 24:1553–1561.

Mitra R, et al. 2013. Functional characterization of piggyBat from the bat Myotis lucifugus unveils an active mammalian DNA transposon. Proc Natl Acad Sci U S A. 110:234–239.

Moran JV, DeBerardinis RJ, Kazazian HH. 1999. Exon shuffling by L1 retrotransposition. Science 283:1530–1534.

Morgante M. 2006. Plant genome organisation and diversity: the year of the junk! Curr Opin Biotechnol. 17:168–173.

Morgante M, et al. 2005. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet. 37:997–1002.

Nosil P, Schluter D. 2011. The genes underlying the process of speciation. Trends Ecol Evol. 26:160–167.

Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. Bioessays 31:703–714.

Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 17:422–432.

Pace JK, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A 105:17023–17028.

Pagán HJ, et al. 2012. Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vespertilionid bats. Genome Biol. 4:575–585.

Pagan HUT, Smith JD, Hubley RM, Ray DA. 2010. PiggyBac-ing on a primate genome: novel elements, recent activity and horizontal transfer. Genome Biol Evol. 2:293–303.

Petrov DA, Hartl DL. 1998. High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. Mol Biol Evol. 15: 293–302.

Phillips CD, Butler B, Fondón JW, Mantilla-Meluk H, Baker RJ. 2013. Contrasting evolutionary dynamics of the developmental regulator PAx9, among bats, with evidence for a novel post-transcriptional regulatory mechanism. PLoS One 8:e76749.

Phillips CJ, et al. 2014. Dietary and flight energetic adaptations in a salivary gland transcriptome of an insectivorous bat. PLoS One 9:e83512.

Platt RN, et al. 2014. Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. Mol Biol Evol. 31:1536–1545.

Poulter RT, Goodwin TJ, Butler MI. 2003. Vertebrate helitrons and other novel Helitrons. Gene 313:201–212.

Price AL, Jones NC, Pezner PA. 2005. De novo identification of repeat families in large genomes. Bioinformatics 21:1351–1358.

Pritchard EJ. 2009. Transposable elements and factors influencing their success in eukaryotes. J Hered. 100:648–655.

Pritchard EJ, Feschotte C. 2007. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. Proc Natl Acad Sci U S A 104:1895–1900.

Putnam NH, et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317: 86–94.

Ray DA, et al. 2008. Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. Genome Res. 18:717–728.

Ray DA, Pagan HJ, Thompson ML, Stevens RD. 2007. Bats with hATs: evidence for recent DNA transposon activity in genus Myotis. Mol Biol Evol. 24:632–639.

Rebollo R, Horard B, Hubert B, Vieira C. 2010. Jumping genes and epigenetics: towards new species. Gene 454:1–7.

Rebollo R, Romansht M, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 46:21–42.

Rensing SA, et al. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69.

Robertson HM. 2002. Evolution of DNA transposons in eukaryotes. In: Craig NL, Cragie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p. 1093–1110.

Rogers RL, Bedford T, Hartl DL. 2009. Formation and longevity of chimeric and duplicate genes in Drosophila melanogaster. Genetics 181: 313–322.

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. Nat Rev Genet. 13: 505–516.

Saravia N, et al. 2013. Human and viral Golgi anti-apoptotic proteins (GAAPs) oligomerize via different mechanisms and monomeric GAAP inhibits apoptosis and modulates calcium. J Biol Chem. 288: 13057–13067.

Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 25:537–546.

Seim I, et al. 2013. Genome analysis reveals insights into physiology and longevity of the Brandt’s bat Myotis brandti. Nat Commun. 4: 2212.

Sela N, Stern A, Makalowski W, Pupko T, Ast G. 2008. Transduplication resulted in the incorporation of two protein-coding sequences into the turmoil-1 transposable element of C. elegans. Biol Direct. 3:41.

Simmons NB. 2005. Order Chiroptera. In: Wilson DE, Reeder DM, editors. Mammal species of the World: a taxonomic and geographic reference. Baltimore (MD): John Hopkins University Press. p. 312–529.

Smith AFA, Hubley R, Green P 1996–2010., RepeatMasker Open-3.0. [cited 2014 Sept 23] Available from: http://www.repeatmasker.org.

Stadelmann S, Lin UK, Kunz TH, Ruedi M. 2007. Molecular phylogeny of New World Myots (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. Mol Phylogenet Evol. 43: 32–48.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.

Sweredoski M, DeRose-Wilson L, Gaut BS. 2008. A comparative computational analysis of nonautonomous Helitron elements between maize and rice. BMC Genomics 9:467.

Tempel S, Nicolas J, El Amriani A, Couee I. 2007. Model-based identification of Helitrons results in a new classification of their families in Arabidopsis thaliana. Gene 403:18–28.

Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. Genome Biol Evol. 2: 656–664.

Thomas J, Sorourian M, Ray D, Baker RJ, Pritham EJ. 2011. The limited distribution of Helitrons to vesper bats supports horizontal transfer. Gene 474:52–58.

Thomas J, Vadnagara K, Pritham EJ. 2014. DINE-1, the highest copy number repeats in Drosophila melanogaster are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). Mob DNA. 5:18.

Wang Q, Dooner HK. 2006. Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc Natl Acad Sci U S A. 103:17644–17649.

Wang W, et al. 2006. High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell 18:1791–1802.

Wicker T, Buchmann JP, Keller B. 2010. Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Res. 20: 1229–1237.

Wolf JB, Lindell J, Backström N. 2010. Speciation genetics: current status and evolving approaches. Philos Trans R Soc Lond B Biol Sci. 365: 1717–1723.

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319:1527–1530.

Xing J, et al. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci U S A. 103: 17608–17613.

Xu JH, Messing J. 2006. Maize haplotype with a helitron-amplified cytidine deaminase gene copy. BMC Genet. 7:52.

Yang LX, Bennetzen JL. 2009a. Distribution, diversity, evolution, and survival of Helitrons in the maize genome. Proc Natl Acad Sci U S A. 106: 19922–19927.

Yang LX, Bennetzen JL. 2009b. Structure-based discovery and description of plant and animal Helitrons. Proc Natl Acad Sci U S A. 106: 12832–12837.

Yang S, et al. 2008. Repetitive element-mediated recombination as a mechanism for new gene origination in Drosophila. PLoS Genet. 4:e3.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 15: 568–573.

Yi SJ, Ellsworth DL, Li WH. 2002. Slow molecular clocks in Old World monkeys, apes, and humans. Mol Biol Evol. 19:2191–2198.

Zabala G, Vodkin LO. 2005. The wp mutation of Glycine max carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell 17:2619–2632.

Zeh DW, Zeh JA, Ishida Y. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. Bioessays 31:715–726.

Zhang G, et al. 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science 339: 456–460.

Zhang JZ. 2003. Evolution by gene duplication: an update. Trends Ecol Evol. 18:292–298.

Zhuo X, Rho M, Feschotte C. 2013. Genome-wide characterization of endogenous retroviruses in the bat Myotis lucifugus reveals recent and diverse infections. J Virol. 87:8493–8501.

Zingler N, et al. 2005. Analysis of 5’ junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5’-end attachment requiring microhomology-mediated end-joining. Genome Res. 15: 780–789.

Associate editor: Richard Cordaux