# INSIGHTS INTO THE TRANSMISSION OF *HELITRONS* AND THEIR IMPACT ON THE GENOME ARCHITECTURE OF *MYOTIS LUCIFUGUS*, THE LITTLE BROWN BAT

by

JAINY THOMAS

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2010

I dedicate my dissertation to my loving husband

Tharun Jose Puthenkandom

ACKNOWLEDGEMENTS

First and foremost I want to thank my advisor Dr. Ellen J. Pritham. I am extremely grateful for all the support, guidance, and encouragement that she has provided throughout my Ph.D. which made my life at UTA, a productive and stimulating experience. The joy and enthusiasm she has for her research was contagious and motivational for me, even during tough times in the Ph.D pursuit. In addition, she was always accessible and always encouraged to bring the best out of me. Altogether, it was a great learning experience and I am really thankful for all her support and financial assistance she provided that helped me to finish my Ph.D in a timely manner.

The members of my Ph.D. committee have contributed immensely to my personal and professional growth at UTA. I am really thankful to Dr. Cedric Feschotte for his stimulating discussions both in and out of the classroom and for all his input and ideas regarding my research. I would like to thank Drs. Esther Betran, Jeff Demuth and Paul Chippindale for kindly serving as my committee members and broadening my perspective on research ideas. I would also like to thank the members of the Mobile DNA and Genome Biology Groups for their comments and input throughout my project. I would like to thank Drs. Shawn Christensen and Elena de la casa Esperon for allowing me to use their labs and equipment.

In my daily work I have been blessed with a friendly and cheerful group of fellow lab mates. I am very thankful to Dr. Cheng Sun for his help with scripts, computer assistance and troubleshooting. Dr. Sarah Shack was always available to give comments on the research and working with her was a great experience. Assiatu Barrie, Claudia Marquez and Komal Vadnagara are great friends and were always helpful and gave me constant support and motivation. They all inspired me in research and life through our interactions during the long

hours in the lab. A special thanks for Jillian North, Megha Bajaj, Rebecca Hornung, Gyanu Kushawah and Mahmuda Farha for being such a wonderful labmates.

I am very thankful to Qi Wang for all his computer assistance and troubleshooting and Dr. Clement Gilbert for his constructive discussions. I would like to thank all my friends Mahima Varma, Mehran Sorourian, John Pace, Mansi Kunte, Eldon Prince, Christi Hull, Blaine Thompson, Jyotiska Chaudhuri, Manish Parihar, Heath Blackmon and all other fellow graduate students who helped me in my life at UTA.

I would also like to thank the Biology Office staff for their help in every need and the Biology Department for providing me with a teaching assistantship.

I would like to thank my family for all their love and encouragement, my parents, C.V Thomas and Mary Thomas and my sister Simi Thomas for their constant love, support, guidance and prayers. Most of all, I am really grateful and thankful to my loving, supportive, encouraging, and patient husband Tharun Jose Puthenkandom whose faithful support during my Ph.D. time was a great blessing and helped me to stay on track. I am very lucky to have such a wonderful family, trusting me always in whatever I do and encouraged me to pursue my dreams.

November 19, 2010

ABSTRACT

INSIGHTS INTO THE TRANSMISSION OF *HELITRONS* AND THEIR IMPACT ON THE GENOME ARCHITECTURE OF *MYOTIS LUCIFUGUS*, THE LITTLE BROWN BAT

Jainy Thomas, PhD

The University of Texas at Arlington, 2010

Supervising Professor: Ellen J. Pritham

Probably the most dynamic proportion of multicellular eukaryotic genomes is composed of teeming populations of parasitic mobile DNA, called transposable elements (TEs). *Helitrons* are rolling circle TEs that have a limited distribution among mammals, having been identified only in the genome of the little brown bat, *Myotis lucifugus*, and a few closely related species. One family of *Helitrons*, called *Helibat*, was estimated to make up as much as 3% of the *M. lucifugus* genome. In addition, *Helibat* was shown to have captured and amplified the promoter and 5' UTR of a highly conserved single copy gene to >1000 copies. Together these observations led to the hypothesis that *Helitrons* have profoundly shaped the evolutionary trajectory of *M. lucifugus* and formed the impetus for this work. We sought to address questions concerning the distribution of *Helitrons* among bats, the role of horizontal transfer (HT) in explaining their patchy distribution and their impact in the *M. lucifugus* genome. To this end, we employed a combination of *in silico*, PCR and DNA hybridization based approaches. We provide for the first time evidence for HT of *Helitrons* (Appendix A). Our analyses revealed a family of *Helitrons* found in *M. lucifugus* as well as an array of distantly related animals, including reptiles, fish, invertebrates, and insect viruses. Most of the HT *Helitrons* were

identified in insects and led us to speculate that the abundance of insects eaten on a daily basis by insectivorous bats might in part influence their propensity for HT. To investigate this hypothesis and to determine the presence of protein coding *Helitrons* in other bat genomes, we examined 83 bat genomes representing ten families with diverse eating habits. These analyses revealed that protein coding *Helitrons* could only be identified in vesper bats and were not detected in a broad range of phyllostomid bats with diverse diets or species representative of any of the other families of bats (Chapter 2). These results suggest that feeding habits alone are not sufficient to explain HT. Finally, we executed a comprehesive analysis of *Helitrons* in the *M. lucifugus* genome (7X coverage) (Chapter 3). This analysis revealed 37 families and 59 subfamilies that contribute to a total of 11.5% of the genome. This is the highest percentage of *Helitrons* ever described in any genome. Through this analysis we show that *Helitrons* have captured promoters, 5' UTRs, 3' UTRs, coding exons and introns of several genes that are well conserved in mammals. These *Helitrons* were further amplified to thousands of copies in some cases. In addition, *Helitrons* have mediated the amplification of several retrogenes. *Helitrons* through HT and amplification have profoundly impacted the genomic architecture of vesper bats and it is tempting to speculate that they tremendously influenced their evolutionary trajectory.

TABLE OF CONTENTS

APPENDIX

LIST OF ILLUSTRATIONS

Figures Page

LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1 Transposable elements and genome evolution

Transposable elements (TEs) are pieces of DNA that can move from one location to another within the genome. They constitute a significant portion of many eukaryotic genomes and often account for the genome size variation observed among closely related organisms (see review Kidwell 2002, Hawkins *et al.* 2006, Piegu *et al.* 2006). Even though TEs are considered as selfish, parasitic, DNA (Orgel and Crick 1980, Doolittle and Sapeinza 1980), recently, the important role of TEs in driving genome evolution has been documented to a great extent (see reviews Feschotte and Pritham 2007, Wessler 2004, Beimont and Viera 2006, Volff 2006, Kazazian 2004, Deninger *et al.* 2003). TEs influence and shape the host-genome in myriad ways, mainly through insertional mutagenesis, chromosomal rearrangements and by providing the raw material that could be manipulated for new gene functions (see review Feschotte and Pritham 2007).

Through insertional mutagenesis, TEs can alter expression of genes in both subtle and substantial ways, from creating allelic diversity to silencing gene or genic regions (see reviews Feschotte and Pritham 2007, Kazazian 1999, Deninger and Batzer 1999, Chen *et al.* 2005). TEs also promote chromosomal rearrangements through ectopic recombination, which may result in gene or segmental duplication, inversions or deletions (Lim and Simmons 1994, Gray 2000). TEs also have influenced the evolutionary trajectory of organisms by providing raw material for the evolution of novel regulatory networks (see review Feschotte and Pritham 2007, Feschotte 2008, Cordaux and Batzer 2009). Domestication of TEs by the host has also led to the birth of many genes (see review Feschotte and Pritham 2007, Lander *et al.* 2001) by contributing regulatory regions, intron splice sites and/or exons (see review Oliver and Greene,

Piryapongsa *et al.* 2007, Nekrutengo and Li 2001, Britten 2006, Bowen and Jordan 2007 and Sorek *et al.* 2002). In addition, the TE encoded machinery can promote retrotransposition of genes, which can evolve to attain beneficial functions (e.g. see review Long *et al.* 2003, Marques *et al.* 2005). Transduction of gene fragments and further amplification could lead to dispersal of those genic fragments across the genome and could lead to formation of novel genetic units (see review Oliver and Greene 2009, e.g. Moran *et al.* 1999, Morgante *et al.* 2005, Jiang *et al.* 2004). Several examples demonstrate that TEs play a vital role in altering host-genome landscape, providing plasticity to the genome and tremendously impacting the evolutionary trajectory of organisms (see review Feschotte and Pritham 2007).

1.2 Horizontal transfer: role in perpetuation of transposable elements and its detection

Vertical transfer (VT) is the transmission of genetic material from parent to offspring where as horizontal transfer (HT) is the transfer of genetic material between reproductively isolated species. HT is fundamental to the evolution of prokaryotes (Ochman *et al.* 2000); in eukaryotes, HT has been shown in the case of genes (for review, see Anderson 2005, Keeling and Palmer 2008) and, more recently there have been cases reporting HT of TEs (e.g., Kidwell 1992, Silva *et al.* 2004, Diao *et al.* 2006, Casse *et al.* 2006, de Boer *et al.* 2007, Loreto *et al.* 2008, Pace *et al.* 2008, Bartolome *et al.* 2009, Roulin *et al.* 2009). Even though, TEs can proliferate in the genome, they typically are inactivated by host silencing machinery and overtime due to the accumulation of mutations (see reviews Hartl 1997). Hence, HT has been proposed as an essential part of the lifecycle of TEs in order to avoid co-evolved host suppression mechanisms aimed at limiting their mobility within lineages (Hart *et al.* 1997, Silva *et al.* 2004, Schaack *et al.* 2010). Even though several vectors including viruses (Marquez and Pritham 2010, Thomas *et al.* 2010) and parasites (Houck *et al.* 1991, Yoshiyama *et al.* 2001, Gilbert *et al.* 2010, Thomas *et al.* 2010) have been proposed to be the shuttles for TEs, the precise mechanisms of HT in eukaryotes remain elusive.

Multiple lines of evidences are used for inferring HT of TEs, including high sequence identity of elements among distantly related species, discontinuous distribution of the elements

across species, and incongruence between the TE phylogeny and host phylogeny (see review Loreto *et al.* 2008). Since TEs are not under host selective constraints and they tend to evolve neutrally, vertical transfer is an untenable explanation for the high sequence identity sometimes observed among TEs identified in distantly related organisms (Pace *et al.* 2008). Stochastic losses or elevated evolutionary rate of TEs in certain lineages can lead to a patchy distribution of the element in closely related organisms and therefore makes this a weak argument for HT in the absence of greater than expected sequence identity (Loretto *et al.* 2008).

### 1.3 Classification of TEs

TEs are classified into class 1 or retrotransposons and class 2 or DNA transposons based on the transposition intermediate (Craig 2002). Class 1 retrotransposons are further classified into three major subclasses, Long Terminal Repeat (LTR) retrotransposons, non-LTRs and DIRS. Retrotransposons transpose through a copy and paste mechanism and utilize an RNA intermediate. DNA transposons transpose by a DNA intermediate and are further divided to three subclasses. Subclass 1 is the classic cut-and-paste DNA transposons, subclass 2 consists of *Helitrons* and subclass 3 is *Mavericks*. Classic cut and paste transposons are characterized by transposase and 12 super families are recognized based on their structural characteristics (Feschotte and Pritham 2007, Bao et al. 2009). *Mavericks* are proposed to transpose by a copy and paste mechanism involving a self encoded DNA polymerase. The classic cut-and-paste DNA transposons and Mavericks have terminal repeats flanking the element and induce target site duplication upon transposition (see review Feschotte and Pritham 2007). *Helitrons* are distinct from other TEs in many ways and are described in detail below.

### 1.4 *Helitrons* and their structural characteristics and implications in genome evolution

*Helitrons* were recently discovered through computational analysis in *Arabidopsis thaliana*, rice *Oryza sativa* and the nematode *Caenorhabiditis elegans*. *Helitrons* constitute about 2% of *A. thaliana* and the *C. elegans* (Kapitonov and Jurka 2001). *Helitrons*, unlike most other DNA transposons which use transposase, encode a putative protein with a rolling circle

initiator motif and PIF1-like DNA helicase domains (Kapitonov and Jurka 2001). Homology of the proteins encoded by *Helitrons* with the single-stranded bacteriophages, plant Gemini viruses, and other bacterial plasmids which undergo rolling circle replication supports the hypothesis that this mechanism facilitates transposition of *Helitrons* (Kapitonov and Jurka 2001, Feschotte and Wessler 2001). *Helitrons*, also known as rolling-circle transposons are different from other DNA transposons in that they lack terminal inverted repeats and do not produce any target site duplication upon transposition. *Helitrons* have atypical structural characteristics with 5' TC and 3' CTRR and a 16-20 nucleotide palindrome 11 nucleotide upstream of the element (Figure 1.1) and get inserted between A and T nucleotides (Feschotte and Wessler 2001, Kapitonov and Jurka 2001).


Figure1.1 Proposed structure of an autonomous animal *Helitron.* Proposed open reading frame consists of Zinc finger domain, a replication initiator motif (REP) and Helicase domains. *Helitron* have a TC at the 5' end and CTRR at the 3' end and a short 16-20 nucleotide palindrome about 11 nucleotide upstream of the element.

*Helitrons* are notorious for their ability to capture and amplify gene fragments. Because of the ability to seize and recombine exons from multiple genes to create novel genetic units, *Helitrons* function as "exon shuffling machines" (Feschotte and Wessler 2001). Even though there are numerous examples of gene capture by *Helitrons* in different organisms, the mechanism of gene capture remains unknown in part because we still do not fully understand how *Helitrons* replicate. In maize, it is estimated that at least 20,000 gene fragments have been picked up and shuffled by *Helitrons* (Morgante *et al.* 2005, Du *et al.* 2009, Feschotte and Pritham 2009, Yang and Bennetzen 2009b, Lal *et al.* 2003, Brunner *et al.* 2005, Gupta *et al.* 2005, Lal and Hannah 2005, Morgante *et al.* 2005, Xu and Messing 2006). Even though, the capture of gene fragments has been reported in many organisms (Cultrone *et al.* 2007, Holister

and Gaut 2007, Sweredoski *et al.* 2008, Choi *et al.*2007, Hoshino *et al.* 2007, Langdon *et al.* 2009, Pritham and Feschotte 2007), this ability is particularly pronounced in maize (Sweredoski *et al.* 2008). The activity of *Helitrons* in maize resulted in intra-species diversity and in the loss of genic co-linearity (Lai *et al.* 2005). The captured gene fragments are sometimes transcribed resulting in chimeric or mosaic transcripts and could impact gene expression in several ways (see review Lal *et al.* 2009). Thus, *Helitrons* are known to create tremendous diversity and dynamically impact the evolutionary trajectory of the host genome.

### 1.5 Distribution of *Helitrons* across eukaryotic tree of life

*Helitrons* have a disparate distribution across the eukaryotic tree of life and have been described in a wide array of eukaryotes including fungi (Poulter *et al.* 2003, Cultrone *et al.* 2007), Oomycetes (Haas *et al.* 2009), plants (Rensing *et al.* 2008, Kapitonov and Jurka, 2001, Lal *et al.* 2003, Yang and Bennetzen 2009a), insects (Kapitonov and Jurka, 2001, Poulter *et al.* 2003, The International Aphid Genomics Consortium 2010, Yang and Bennetzen 2009a, Langdon *et al.* 2009), nematodes (Kapitonov and Jurka 2001), vertebrates (Poulter *et al.* 2003, Zhou *et al.* 2006) and mammals (Pritham and Feschotte 2007). In some cases, *Helitrons* constitute a significant portion of the genomes, e.g. 2-3% in *C. elegans*, *A. thaliana* (Kapitonov and Jurka 2001) and *Zea mays* (Yang and Bennetzen 2009) and 5% in *Drosophila virilis* (Kapitonov and Jurka 2007), the highest contribution by *Helitrons* reported to date. Interestingly, *Helitrons* have a unique distribution across mammals. The presence of a single family of *Helitrons* was reported in the genome of *Myotis lucifugus*, the little brown bat. The putative autonomous family of *Helitrons* named as *Helibat1* and the subsequently amplified non-autonomous families together constituted almost 3.4% of the *M. lucifugus* genome (Pritham and Feschotte 2007). No evidence for the presence of *Helitrons* was identified in the 44+ mammalian genomes sequences available in the public database (Figure 1.2; Pritham and Feschotte 2007). *Helitrons* were detected in the genome of three other bats which diverged from *M. lucifugus* around 16-25 mya and all of these bats belong to Vespertilionidae family. Based on the sequence divergence data, it was estimated that *Helitrons* became active in the

lineage of bats around 30-36 mya (Pritham and Feschotte 2007). Degenerated remnants of *Helitrons* were found in the genome of Platypus (*Ornithorhynchus anatinus*) suggesting that *Helitrons* might have colonized the ancestral mammalian species, which diverged from the present day mammals around 180mya (Figure 1.2). The two possible explanations for the presence of *Helitrons* are: either these *Helitrons* might have been preserved by intense purifying selection and became active in the lineage of *M. lucifugus* or it was horizontally introduced into the genome. Since TEs are not under host-selective constraints, HT is the most parsimonious explanation for the presence of *Helitrons* in the *M. lucifugus* genome and its absence in other mammalian genomes. Even though HT has also been invoked to occasionally explain the sequence similarity and discordant distributions observed across highly diverged taxa, there are no convincing evidences supporting the HT of *Helitrons*. Bats could be an interesting organism to look for this evidence, since they are the only group of mammals that are known to harbor *Helitrons* (Pritham and Feschotte 2007).

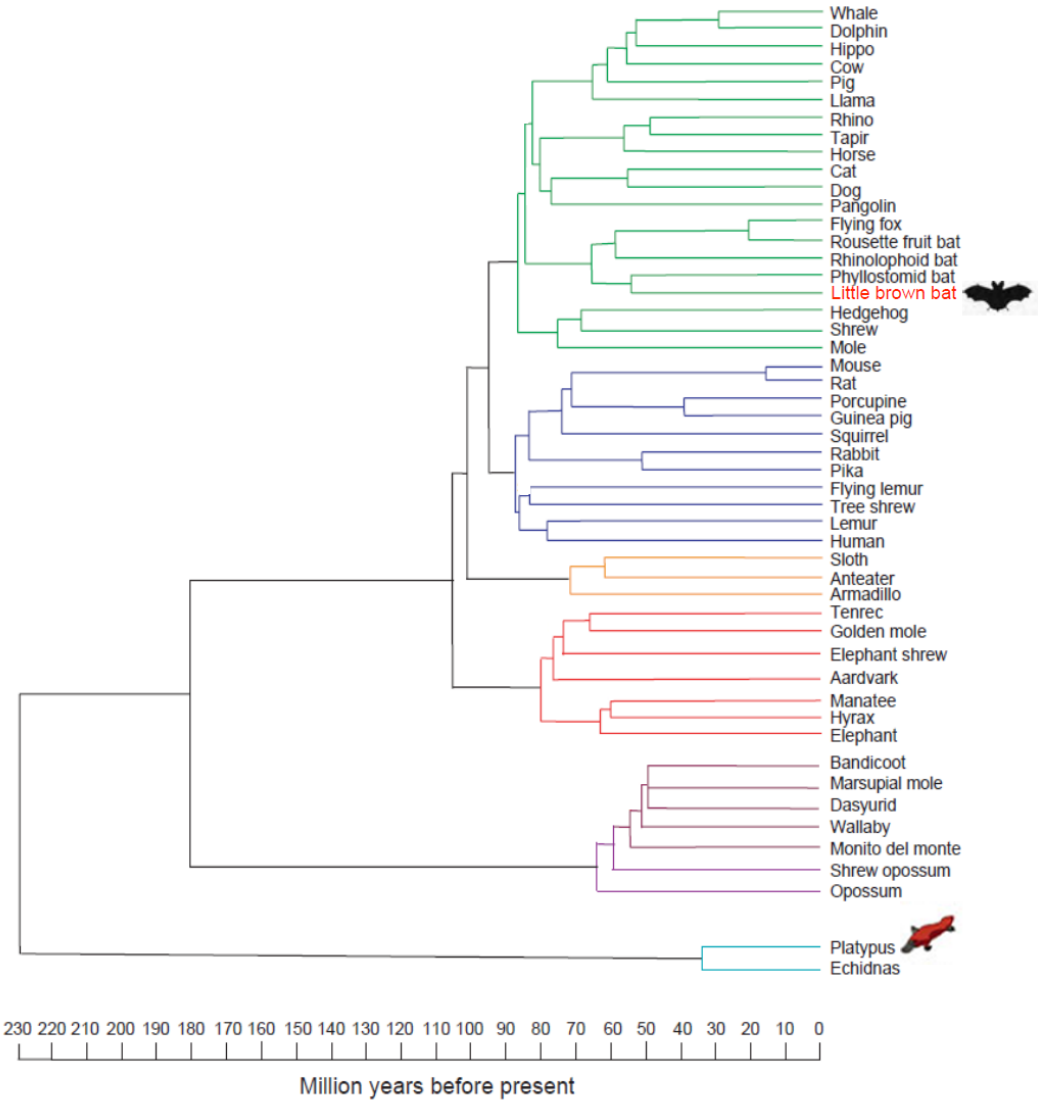



Figure1.2. Phylogenetic distribution of *Helitrons* in mammals. Presence of *Helitrons* are found only in little brown bat (*Myotis lucifugus*) and platypus (*Ornithorhynchus anatinus*) among 44+ mammalian genome sequences available. Tree was modified from (Murphy *et al.* 2004)

## 1.6 Bats, the remarkable mammals

Bats are 'remarkable mammals' that constitute more than 20% of extant mammals (Order Chiriptera; ≈1100 species; Simmons 2005, Calisher *et al.* 2006). They are the only group of mammals that have attained true and sustained flight. In addition to their unique ability to fly, they echolocate and also have tremendous diversity in their diet ranging from plants, insects and vertebrates to blood. They are most geographically dispersed mammal other than human

(Nowack 1994). Chiropterans also play an important role as seed dispersers, pollinators (750 plant species; Neuweiler 2000) and pest controllers in the eco-system (Williams-Guillén *et al.* 2008, Kalka *et al.*2008). It is estimated that one million bats could eat about 10,000 kg of insects every night (McCracken 1996) and recent studies demonstrate the role played by bats in significantly reducing the arthropod abundance and there by indirectly reducing the herbivory (Kalka et al 2008, Williams-Guillén *et al.* 2008); sometimes, greater than that of birds. However, they are also natural reservoirs of many deadly viruses including severe acute respiratory syndrome (SARS), Corona, Rabies, and Ebola that have been shown to have potential implications in human and veterinary health (see review Calisher *et al.* 2006). Traditionally bats are classified as megabats (fruit bat) and microbats (Nowack 1994). Recently, based on more elaborative study, the order Chiroptera is divided into two sub-orders Yinpterochiroptera and Yangochiroptera (Fig 1.3). Yinpterochiroptera constitute Pterpodidae and Rhinolophoidea (four families) whereas Yangochiroptera constitute Emballonuroidea, Noctilionoidea, Vesepertilinoidea superfamilies (13 families; Teeling *et al.* 2005, Miller-Butterworth *et al.* 2007). Among the different families, the family Vespertilionidae has 407 species and is the most diverged family, while the family Phyllostomidae has 160 species and the family Molossidae has 100 species (see review Calisher *et al.* 2006).

## 1.7. Significance and aims of the study

My interest in *Helitrons* in the *M. lucifugus* genome began with the discovery of one lineage of *Helitrons* called *Helibat* in the 2x coverage of the *M. lucifugus* genome sequence (Pritham & Feschotte 2007). Although, *Helitron* fragments were detected in some orthologous positions in partial sequence data available for three vesper bats closely related to *M. lucifugus* (Pritham and Feschotte 2007) no *Helitrons* could be identified in any of the other 44+ sequenced mammalian genomes, representing a great diversity of lineages (Figure 1.3). This study estimated that *Helibats* have amplified to more than 100,000 copies constituting ~3.4% of the *M. lucifugus* genome. Based on the sequence divergence data, the estimated age of activity of this family of *Helitrons* was 30-36 mya (Pritham and Feschotte 2007). More precisely, the

presence of *Helitrons* was also detected in *M. myotis*, *Kerivoula papilosa* and *Pipistrellus abramus* (Pritham and Feschotte 2007). Diversification of vesperbats (16-25 mya; Teeling *et al.* 2005) is roughly coincidental with the period of activity of *Helitrons* (30-36mya; Pritham and Feschotte 2007). Understanding the distributions of *Helitrons* across the major families of bat and estimating the number of independent invasions may help us in appreciating the nature and impact of *Helitron* amplification. We hypothesized that *Helitrons* might have invaded other bat families. To test this hypothesis, we used a combination of *in silico*, PCR and DNA hybridization based approaches to detect the presence of *Helitrons* across bat families (Chapter 2). We ask the following questions: What is the distribution of *Helitrons* across different bat species? How many independent invasions of *Helitrons* in different bat lineages can we detect?

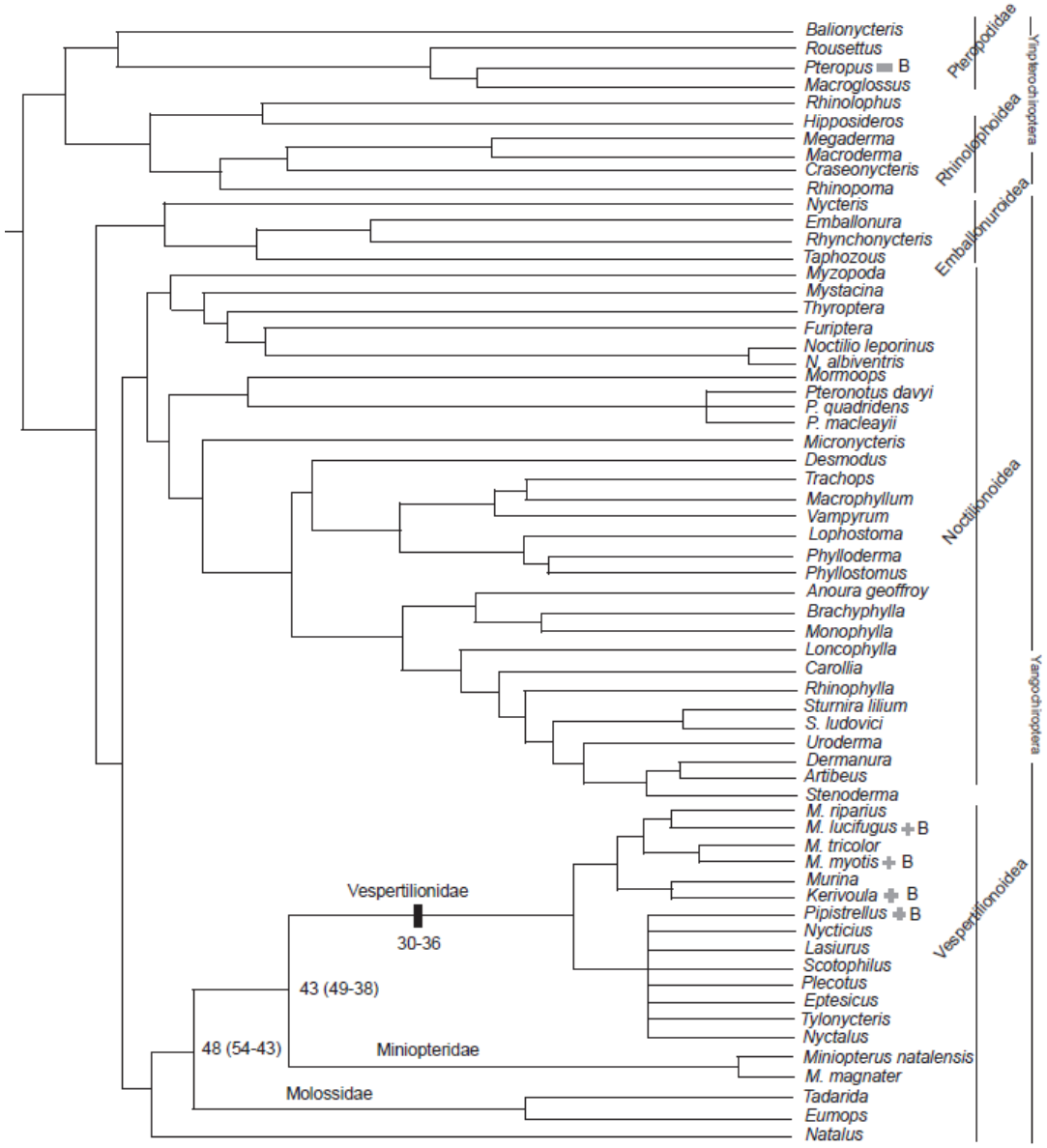



Figure 1.3. Phylogenetic distribution of *Helitrons* in bats. (+) indicates the presence of *Helitrons* and (-) denotes the absence of *Helitrons* and (▮) shows the estimated age of amplification (in million years, my) of *Helibats* in the *Myotis lucifugus* genome (Pritham and Feshotte 2007) and the numbers at the nodes are the molecular dates in millions of years, values in parentheses are the 95% credibility intervals. The letter B indicates the method used in determining the distribution of *Helitrons* in different bat species [B- bioinformatic analysis (blastn analysis against nucleotide collection (nr/nt) database or whole genome shot gun sequences)] (Pritham and Feshotte 2007). (Tree was redrawn using Teeling *et al.* 2005, Stadelman *et al.* 2007, Hoffmann *et al.* 2008, personal communication with Baker R.J and divergence data was obtained from Miller-Butterworth *et al.* 2007)

The distinctive presence of *Helitrons* in *M. lucifugus* due to intense conservation of these genomic parasites and absence of these elements in other 44+ mammalian lineages (including other bat families) due to stochastic losses is untenable by vertical inheritance. The most parsimonious explanation for this scenario is the HT. Despite some reports of HT of some *hAT* TEs in mammals (Pace *et al.* 2008; Gilbert *et al.* 2010), no cases of HT of *Helitrons* have ever been described. The unique presence of *Helitrons* in the *M. lucifugus* genome among mammals led us to hypothesize that HT was responsible for the presence of *Helitrons* in the *M. lucifugus* genome. To test this hypothesis we systematically analyzed the *M. lucifugus* genome and all the other genome sequences that are publicly available for the presence of related *Helitron* families (Chapter 4 as Appendix A). We answer the following questions: How many and what kind of *Helitrons* are present in the 7X *M. lucifugus* genome? Do we find any additional evidence supporting HT of *Helitrons* in the *M. lucifugus* genome? Can we identify any potential mechanisms behind the horizontal transmission?

In addition to contributing to a significant portion (~3%) of the genome, *Helitrons* have captured and amplified a gene fragment in the *M. lucifugus* genome. *Helitrons* have transduced the promoter, 5' UTR and the first protein coding exon of the gene NUBPL (nucleotide binding protein like; NUBPL is a single-copy gene encoding a highly conserved protein in mammals). *Helitron* along with the captured fragment has amplified to >1,000 copies, in the little brown bat genome. It is intriguing to speculate that gene capture is a general phenomenon of *Helitrons* in *M. lucifugus* that has contributed to the evolution of the genetic architecture of this species. We hypothesize that *Helitrons* have captured, shuffled and amplified gene fragments recurrently. To test this hypothesis, we systematically mined all *Helitrons* (>3 copies) from the 7x genome and examined them closely using homology based methods for sequences highly conserved in other mammalian genomes (Chapter 3). We answer the following questions in chapter 3: How many novel *Helitrons* families are present in the *M. lucifugus* genome? How much do *Helitrons* contribute to the *M. lucifugus* genome? Can we identify other examples of gene captures? Do bat *Helitrons* have a similar tendency as that of plant *Helitrons* in capturing gene fragments?

What is the nature and pattern of gene captures in the bat genome? Can we propose any model based on the gene capture data? What could be the possible impact of the gene captures?

The main research questions of the study can be summarized as three questions listed below.

Research questions:

1. What is the distribution of *Helitrons* in bats?

2. Is there any HT of *Helitrons* to the *M. lucifugus* genome?

3. What is the impact of *Helitrons* in the genome of *M. lucifugus*?

# CHAPTER 2

# DISTRIBUTION OF *HELITRONS* IN DIFFERENT SPECIES OF BATS

## 2.1. Introduction

Transposable elements (TEs) are mobile pieces of DNA that can move from one location to another within the genome. They constitute a significant portion of many eukaryotic genomes and are considered as selfish and parasitic DNA (for review Feschotte and Pritham 2007). Due both in part to their abundance as well as their mutagenic properties, TEs can alter the host-genomic landscape and can play a dynamic role in genome evolution (for review Feschotte and Pritham 2007, Feschotte 2008, Oliver and Green 2009, Pritham 2009, Zeh *et al.* 2009).

Based on the transposition intermediate, TEs are classified as class 1, retrotransposons and class 2, DNA transposons (Craig 2002). *Helitrons* are distinct from other DNA transposons in that they lack terminal inverted repeats and do not produce any target site duplications. Also known as rolling-circle transposons, these elements have atypical structural characteristics with 5' TC and 3' CTRR. Based on their protein homology with bacterial plasmids and other viruses that replicate by rolling circle mechanism, it has been proposed that *Helitrons* transpose in a similar fashion (Kapitonov and Jurka 2001, Feschotte and Wessler 2001). *Helitrons* are also known for their propensity to capture host gene sequences and their ability to recombine them to produce novel genetic units (e.g. Morgante *et al.* 2005).

Even though TEs are long standing residents of eukaryotic genomes; there is extreme diversity in TE composition (see review Feschotte and Pritham 2007, Pritham 2009). In mammals, ~50% of the genome is typically made up of TEs, most of which are retrotransposons. Even though, cut-and-paste DNA transposons also constitute a significant

proportion (≈3%) of the human genome, there was no evidence for recent activity (up to 40 mya; Lander *et al.* 2001, Waterson *et al.* 2002, Pace and Feschotte 2007). Nonetheless, there have been reports of a wide array of recent cut-and-paste DNA transposon activity (Ray *et al.* 2007, Ray *et al.* 2008), including *Helitrons* (Pritham and Feschotte 2007), in the genome of the little brown bat (*Myotis lucifugus*). Amplification of the *Helibats* preceded the activity of other DNA transposons in the bat genome (Ray *et al.* 2008). The putative autonomous family of *Helitrons* (*Helibat1*) and the subsequently amplified non-autonomous families together constituted almost 3.4% of the *M. lucifugus* genome (Pritham and Feschotte 2007).

Degenerate remnants of *Helitrons* are found in the platypus genome, *Ornithorhynchus anatinus*, which indicates the colonization of a family of *Helitrons* in the ancestral mammalian or monotreme species (Pritham and Feschotte 2007). However, no evidence for the presence of *Helitrons* is found in 44+ mammalian genomes including the megabat, *Pteropus vampyrus* (family Pteropodidae; Chiroptera). On the other hand, *Helitrons* were also identified in three other closely related bats at orthologous positions indicating the amplification of *Helitrons* occurred prior to the split of those species (Pritham and Feschotte 2007). These bats belong to the Vespertilionidae family and are known to be the most speciose family in the order Chiroptera.

Chiropterans constitute more than 20% of extant mammals (≈1100 species; Simmons 2005) and are the most geographically dispersed (Nowack 1994). In addition to their unique ability to fly, they echolocate, and also have tremendous diversity in diet including plants, insects, vertebrates and blood. Chiropterans also play an important role as seed dispersers, pollinators (750 plant species; Neuweiler 2000) and pest controllers in the eco-system (Williams-Guillén *et al.* 2008, Kalka *et al.* 2008). It is estimated that one million bats could eat about 10,000 kg of insects every night (McCracken 1996). The order Chiroptera is comprised of two sub-orders Yinpterochiroptera and Yangochiroptera. Yinpterochiroptera constitute Pterpodidae and Rhinolophoidea (four families) whereas Yangochiroptera constitute

Emballonuroidea, Noctilionoidea, Vesepertilinoidea superfamilies (13 families; Teeling *et al.* 2005, Miller-Butterworth *et al.* 2007).

The remarkable activity of *Helitrons* in the *M. lucifugus* genome, and the discontinuous distribution of these elements among mammals lead us to investigate the extent of colonization of *Helitrons* among the diverse species of bats belonging to different families. We investigated the presence of *Helitrons* in 83 species of bats belonging to ten extant families and we report on the restricted distribution of *Helitrons* to vesper bats among chiropterans.

## 2.2. Methods

To screen a wide phylogenetic sample of bats for *Helitrons*, a degenerate PCR based approach coupled with Southern and slot blot analysis was employed in conjunction with bioinformatic screening of sequences publically available.

### 2.2.1 PCR based screening for *Helitrons*

#### 2.2.1.1 DNA extraction

DNA or tissue samples representing 63 different bat species belonging to ten different families were collected from different sources (Ambrose Monell Cryo Collection at American Museum of Natural History, New York, The Museum of Texas Tech University- Lubbock, Dr. David Ray, Mississippi State University, personal collection; Appendix B). DNA was extracted from tissues preserved in lysis buffer/ ethanol using 'DNeasy Blood and Tissue extraction kit' Spin Column (Qiagen Sciences, Maryland, USA) following the manufacturer's protocol.

#### 2.2.1.2 PCR, Cloning and Sequencing

Degenerate primers were designed based on a unique and conserved 350 bp region, the replication initiator (Rep) motif of the protein encoded by the putative autonomous family of *Helitrons* (Fig 1; forward primer 5' TTYATHACBTTYACVTGYAATCC3' and reverse primer 5' CCACATGGHCCATGTACCAT3'). Polymerase Chain Reaction (PCR) was carried out with degenerate primers with an initial denaturation step of 2 min at 94℃ followed by 35 cycles of 30s at 94℃, 30 s at 47℃, and 30s at 72℃ and a fi nal extension of 10 min at 72℃. PCR mix was: Buffer (1X); MgCl2 (4 mM); dNTP (0.2 mM), Forward primer (1.2 pM), Reverse primer (0.4

pM), BSA (0.1µg/ µl), Taq (1.25 U; GoTaq Flexi DNA polymerase, Promega Corporation, Madison, USA), DNA 30 ng; and $H_2O$, ≤25 µl. PCR products were cloned using Strata Clone PCR cloning kit, (Stratagene, La Jolla, USA) and was sequenced using the AB Big Dye Terminator protocol version 3.1(Applied Biosystems, Inc., Foster City, California). DNA sequences were generated using a 3130xl Genetic Analyzer (Applied Biosystems) sequencer. Sequences were verified using Sequencher version 4.7 (Gene Code Corporation, Ann Arbor, Michigan).

2.2.2 Hybridization based screening of *Helitrons*

Followed by the initial screening for *Helitrons* using PCR, cloning and sequencing, the presence or absence of *Helitrons* were further validated by hybridization based methods, slot blot or Southern blot analysis.

2.2.2.1 Slot blot

2.2.2.1.1 Membrane preparation

To denature the DNA, NaOH (final concentration 0.4M) and EDTA (final concentration 10mM) was added to approximately 1µg of total genomic DNA, and was boiled for 10 minutes. The denatured DNA was loaded sequentially onto a nylon membrane (Amersham Hybond –XL, GE Healthcare, Bucks, UK) using a vacuum manifold. Both to estimate *Helitron* copy number and as a positive control for the hybridization, different amounts of the plasmid DNA (Strataclone PCR cloning vector pSC-A-amp/kan; 4.3kb) containing the *Helitron* fragment (350 bp) was loaded (0.4ng, 1ng, 2.5ng, 10ng, 20ng and 25ng) on the same nylon membrane. In addition, we also used different amounts (0.4ng, 1ng and 2.5ng) of the plasmid DNA (Strataclone PCR cloning vector, pSC-A-amp/kan; 4.3kb) containing a Rag1 (Recombination activation gene; a single-copy gene highly conserved across jawed vertebrates) fragment (387 bp) as a positive control for the hybridization. Rag1 hybridizations were conducted to confirm the presence of DNA on the membrane and to illustrate that elements with low copy number (even single copy) would be identified through this technique. Mouse DNA was also loaded on the membrane as a negative control for the hybridization with the *Helibat* probe as the mouse

genome does not contain any *Helitrons*. The membrane was rinsed in 2XSSC and was crosslinked using an UV crosslinker (Spectroline, Giangarlo Scientific Co, Inc.USA).

2.2.2.1.2 Probe generation and labeling

*Helitron* and Rag1 probes were generated from the plasmid DNA containing the respective PCR amplified product. *Helitron* fragments were PCR amplified from the *Myotis sp*. using the degenerate primers mentioned in the section 2.1.2. The Rag1 gene was amplified from *Myotis* sp. using forward primer 5'ATG GGA GAT GTG AGC GAG AAG CAT3' and reverse primer 5'ACA AAG GGT GCA GAT GTA GAC GGA3'. PCR was carried out with an initial denaturation step of 2 min at 94℃ followed by 30 cycles of 30s at 94℃, 30 s at 57℃, and 30 s at 72℃ and a final extension of 10 min at 72° C. The PCR mix included: Buffer (1X), MgCl2 (3 mM), dNTP (0.2 mM), Forward primer (0.4 pM), Reverse primer (0.4 pM), BSA (0.1µg/ µl), Taq (0.75 U; GoTaq Flexi DNA polymerase, Promega), DNA 30 ng; and $H_2O$, ≤25 µl. Amplified PCR products of both *Helitron* and Rag1 gene were cloned using Strata Clone PCR cloning kit, (Stratagene, La Jolla, USA). The probes were radiolabeled with dCTP-$\alpha^{32}$P (Perkin Elmer, Massachusetts, USA) using random primed labeling kit (Roche, Basel, Switzerland). The labeled probes were purified using QIAquick PCR Purification Kit (Qiagen Sciences, Maryland, USA).

2.2.2.1.3 Hybridization and developing

The membrane was pre-hybridized with denatured salmon sperm DNA (100 µg/ ml) and hybridization solution (6XSSC, 0.5% SDS, 5X Denhardt's solution and 1mM EDTA) for 2 hours. After that, the membrane was hybridized using a fresh hybridization solution, denatured radio-labeled probe and salmon sperm DNA (100 µg/ ml). The membrane was first hybridized with Rag1 probe under high stringency (65℃ for 12 hours ) conditions followed by low stringency (2xSSC, 0.1% SDS at 55℃) to medium stringency (2xS SC 0.1% SDS at 65℃) washes. The membrane was exposed to film (Kodak-Biomax MS, Carestream Health Inc. Rochester, USA) in a cassette and stored at -80℃ overnight or longer as required. Stripping of the Rag 1 probe from the membrane was done using 0.4M NaOH and stripping buffer (0.1%SDS, 0.1XSSC,

0.1MTris HCl at pH 7.2) at 45℃. Hybridization with *Helibat* probe following the same conditions mentioned above was undertaken after the Rag 1 probes were completely stripped off from the membrane.

2.2.2.1.4 Copy number estimation

Slot blot hybridization technique was used to estimate the copy number of the *Helitrons* in the vesper bat genomes. The number of genomes present in the serial dilutions of plasmid was calculated from the amount of DNA loaded (see section 2.2.1.1). The number of genomes present in the vesper bat DNA loaded was also calculated (see section 2.2.1.1). The intensity of the hybridization signal found with the *Helitrons* in the bat genomes and plasmid was compared by eye and the copy number was estimated based on the amount of DNA loaded onto the membrane.

2.2.2.2 Southern blot

Five µg of genomic DNA digested with *Bam* HI enzyme (Promega) at $37^0$C for 16-18 hours from selected bat species and was electrophoresed on 0.8% agarose gel at 20V for 12 hours. The DNA from the gel was transferred to a nylon membrane (Amersham Hybond –XL, GE Healthcare, Bucks, UK) by the alkaline transfer (Brown 1999) method. The membrane was crosslinked using UV crosslinker (Spectroline, Giangarlo Scientific Co, Inc.USA) and was hybridized with the *Helitron* probe (same procedures mentioned in section 2.2.1.2 were used for generating the probe). The membranes were hybridized washed by following the same conditions (section 2.2.1.3) used for the slot blot analysis.

2.2.3 Cytochrome b gene sequencing

The mitochondrial cytochrome b gene was sequenced to confirm the identity of the bat samples. PCR amplification of the cytochrome b gene was completed using the primers and conditions mentioned in Hoffman and Baker (2001). The PCR products were purified using QIAquick PCR Purification Kit (Qiagen Sciences, Maryland, USA) following manufacturer's instructions and sequenced using AB Big Dye Terminator protocol (see section 2.1.2 for details regarding sequencing) using one of the amplification primers. Sequences were blasted (blastn,

default parameters) against the nucleotide collection (nr/nt) database at National Center for Biotechnology Information (NCBI), and best hits were noted (Appendix).

2.2.4 *In silico* identification of *Helitrons*

Blast searches (BlastN default parameters) were performed using *Helitrons* identified in the *M. lucifugus* genome to find *Helitrons* in sequences deposited in the different databases at NCBI including nr/nt, high throughput genome sequences, expressed sequence tags and genome survey sequences. Full length elements or fragments with diagnostic last 30 bp at the 3' end, an evalue of ≥ 3e-06, and at least 200 bp in length were considered as significant hits to the *Helitron* query.

## 2.3. Results

2.3.1 Restricted invasion and amplification of *Helitrons* in the vesper bat lineage and its absence from other bat families

To analyze the distribution of *Helitrons* across the bat lineage and to estimate the number of independent invasions, screening for the presence or absence of the *Helitrons* was conducted in different bats belonging to major families representing each superfamily. Blast searches were undertaken to determine the presence of *Helitrons* in bats with whole genome sequences (little brown bat, *M. lucifugus* and fruit bat, *Pteropus vampyrus*) or any gene sequences deposited in the GenBank. Through blast searches, we found evidence for the presence of *Helitrons* in different vesper genomes including *Eptesicus serotinus* (EU751001.1), *M. tricolor* (EU751022.1), *M. volans* (GU197975.1), *M. thysanodes* (GU197965.1), *M. keeni* (GU197973.1), *M. evotis* (GU197972.1), *M. sodalis* (HQ127377.1), *Scotophilus kuhlii* (EU751015.1), *S. nux* (EU751017.1), *S. leucogaster* (EU751018.1), *S. nigrita* (EU751020.1), *S. marovaza* (EU751021.1), *S. dinganii* (EU751002.1), *S. heathii* (EU751011.1), *S. viridis* (EU751016.1), *Nyctalus leisleri* (GU385818.1), *Tylonycteris pachypus* (EU552462.1) and *Plecotus rafinesquii* (FJ469635.1).

Due to the limited availability of bat genomes with sequencing projects, a PCR and hybridization based strategy was employed to screen DNA from a wide and divergent array of

bats species. Degenerate PCR primers were designed based on the replication initiator (Rep) region of the protein encoded by *Helitrons*, which is highly conserved and unique to this element (Figure 2.1; Pritham and Feschotte 2007). The amplification of diverse families of *Helitrons* (ranging in protein similarity from 52-91%) using these primers suggested that this approach would be effective in uncovering distantly related families from diverse genomes including plants, insects and vertebrates (e.g. see Supplemental Fig. 2.1.1). For this study, DNA encoding *Helitron* protein fragments from 17 different vesper bat species (see Appendix B) were successfully amplified (Appendix D).

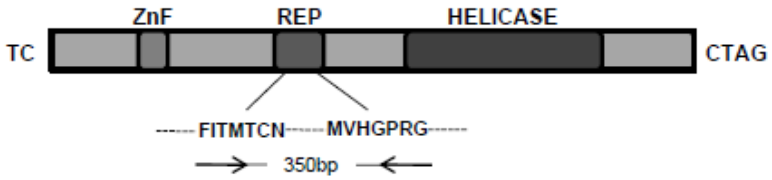



Figure 2.1. Cartoon depiction of an autonomous *Helitron.* The amino acids represent the region to which the degenerate primers were designed. The arrows indicate the 350 bp region amplified to generate the probe for the Southern blot and slot blot hybridizations.

To rule out contamination, we assayed for the presence of autonomous *Helibat* families in vesper bats by using a non-PCR based hybridization (slot blot or Southern blot) technique. The membranes containing DNA samples of different bat species were hybridized with a *Helitron* probe (350 bp, Figure 2.1) at high stringency conditions and were washed at low to medium stringency conditions (elements with >71% sequence identity were hybridized). Using hybridization techniques, *Helitrons* were identified in nine vesper bats (every sample for which we had sufficient DNA), (Figure 2.2, Appendix E). We extrapolated that the estimated copy number of autonomous *Helitrons* for the six species included on the slot blots was between ≈1200 to 5000 (Appendix E). Therefore by employing *in silico*, PCR and hybridization techniques, we show that *Helitrons* are present in 35 vesper bats representing all four subfamilies of Vespertilionidae and that they have amplified to a significant copy number in the six species analyzed using slot blots (Appendix E).

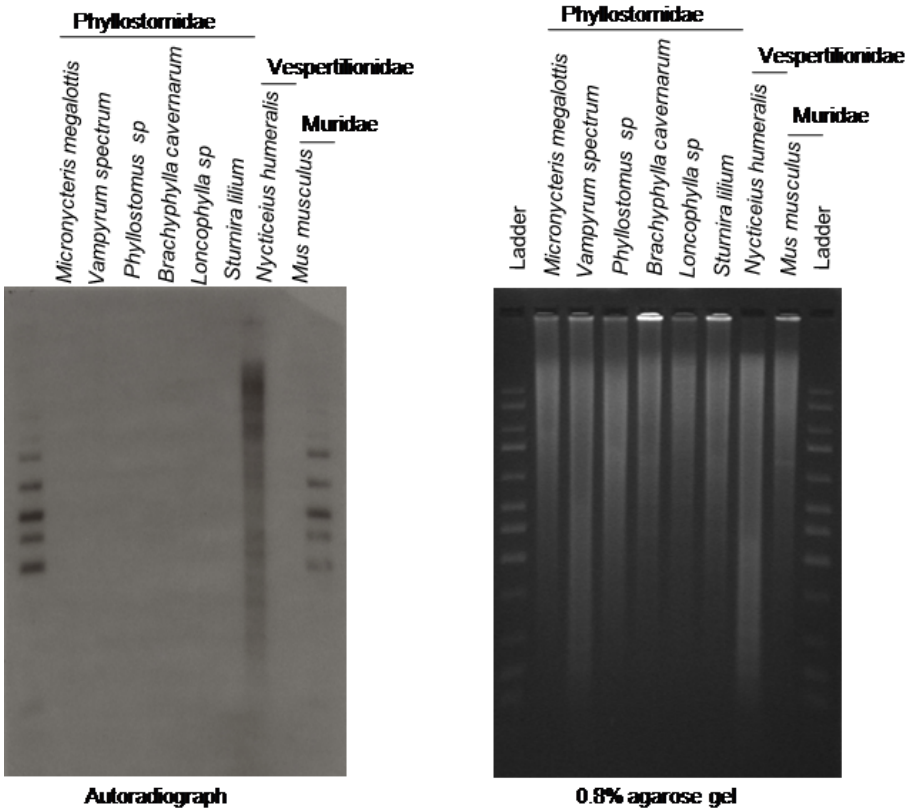



Figure 2.2 Example of genomic Southern blot. a) DNA samples of seven different bat species and mouse (negative control) were cut with *Bam*H1 and were electrophoresed on a 0.8% agarose gel. b) Autoradiograph of the Nylon membrane, blotted with DNA samples after hybridization with the radio-labeled *Helibat* probe.

Our analysis, also included bat species belonging to ten different families representing all superfamilies of Chiroptera, using the same techniques as outlined above. No products could be amplified using degenerate primers from the other 45 bat species belonging to Pteropodidae, Megadermatidae, Emballonuridae, Rhinolophidae, Noctilionidae, Mormoopidae, Phyllostomidae, Thyropteridae, Molossidae and Miniopteridae families (Figure 2.1, Appendix D). In addition, no hybridization signal was detected with the *Helibat* probe in any of 40 bat samples tested, representing those bat families (Pteropodidae, Megadermatidae, Rhinolophidae, Noctilionidae, Mormoopidae, Phyllostomidae, Thyropteridae, Molossidae and Miniopteridae) or in the mouse DNA (negative control) examined using slot blot and Southern

blot (Figure 2.2, Appendix E, Appendix D). As a positive control, the membranes were stripped and rehybridized with a Rag1 probe (387 bp). The hybridization of these bat DNA samples with the Rag1 gene probe confirmed that the absence of signals with the *Helitron* probe was not due to absence of DNA or to low copy number of the elements. In addition, we did not find any evidence for the presence of *Helitrons* in the sister family of the Vespertilionidae, the Miniopteridae (*Miniopterus magnater*) either by PCR or hybridization methods (Figure 2.3, Appendix D, Appendix E). These results further validate that the invasion and amplification of *Helibat* occurred only in the vesper bat lineage. Since the presence of *Helitrons* was not identified in other major bat families, even in the sister family, Miniopteridae, we conclude that *Helitrons* are restricted to the vesper bat lineage.

## 2.4. Discussion

### 2.4.1 Unique distribution of *Helitrons* across bats

Our work demonstrates that *Helitrons* are restricted to vesper bats (family Vespertilionidae) among 83 bat species tested representing all four superfamilies belonging to the two suborders of Chiroptera. The estimated timing of the colonization of *Helitrons* in the vesper bats is 30-36 mya (based on sequence analysis; Pritham and Feschotte 2007) which is compatible with the finding that *Helibats* are not present in the sister family, the Miniopteridae (Figure 2.3, Appendix D, Appendix E) which diverged 43 mya from the vesper bat lineage (Miller-Butterworth *et al.* 2007). This dating is congruent with the identification of *Helitron* insertions at orthologous positions in three vesper bats (*M. myotis*, *Kerivoula papillosa*, and *Pipistrellus abramus*; Pritham and Feschotte 2007). The representatives from ten of the 18 recognized bat families (Teeling *et al.* 2005) do not have *Helitrons* in their genomes, based on the taxa sampled. Such phylogenetic structure presents a unique opportunity to study the evolutionary consequences of intragenomic movement of this transposon in the vespertilionid species as compared to species of other bat families where this *Helitron* is absent. In our sample of 35 vespertilionid species representing 11 genera, all had evidence of *Helitrons* in their genomes. An extension of these results is the hypothesis that most if not all species of

vespertilionids have *Helitrons* and therefore there are hundreds of species that can be examined to understand the impact of *Helitron* activity in this lineage. This *Helitron* invasion does not appear to be a singular event. Indeed, the vesper lineage appears to have become a hub for the colonization and activity of many other DNA transposons, which have invaded in waves at different times ranging from 1.1- 33.3 mya (Ray *et al.* 2008). This pattern of recurrent invasion and colonization of DNA transposons is in striking contrast to many other mammalian lineages where DNA transposon activity has been reported to have ceased almost 40 mya (Pace and Feschotte 2007). The factors favoring the colonization and amplification of *Helitrons* and other DNA transposons in the vesper bat lineage are mysterious and beg for further study.

2.4.2 Role of horizontal transfer and vesper bats

The invasion and colonization of *Helitrons* in vesper bats among mammals supports the acquisition of these elements by horizontal transfer (HT). Indeed, some *Helitrons* and other DNA transposons identified in the genome of the little brown bat have already been shown to have under gone horizontal transmission between distantly related animals including other mammals, reptiles, amphibians, fish, nematodes, insects as well as viruses that infect insects (Thomas *et al.* 2010, Pace *et al.* 2008, Novick *et al.* 2010, Gilbert *et al.* 2010, Pagan *et al.* 2010). The prevalence of cases of HT of TEs in vesper bats and the recurrent and on-going nature of this phenomenon prompts the question; why vesper bats? What is special about this group? Vesper bats are unique among bats in many ways; the totality of the geographic and ecological distribution of vespertilionid bats is the greatest for any non-human mammalian family (Nowack 1994).

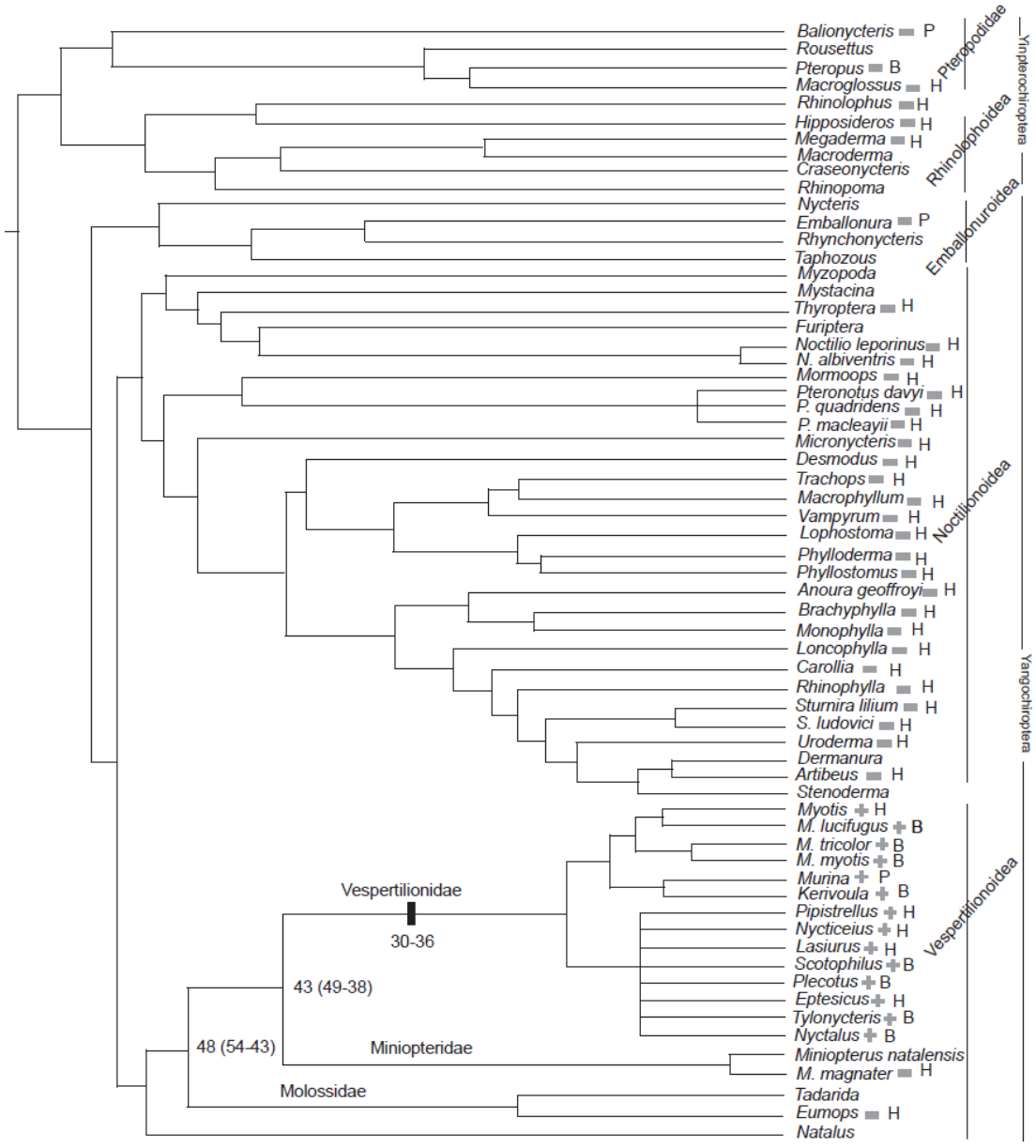



Figure 2.3. Phylogenetic distribution of *Helitrons* in bats. (+) Indicates the presence of *Helitrons* (-) denotes the absence, (▮) shows the estimated age of amplification (in million years, my) of *Helibats* in the *Myotis lucifugus* genome (Pritham and Feshotte 2007) and the numbers at the nodes are the molecular dates in millions of years, values in parentheses are the 95% credibility intervals. The letter P, H and B indicates the method used in determining the distribution of *Helitrons* in different bat species [P- using degenerate PCR, H- slot blot / Southern blot hybridization, B- bioinformatic analysis (blastn analysis against nucleotide collection (nr/nt) database or whole genome shot gun sequences)] (Tree was redrawn using Teeling *et al.* 2005, Stadelman *et al.* 2007, Hoffmann *et al.* 2008, personal communication with Baker R.J and divergence data was obtained from Miller-Butterworth *et al.* 2007)

Further the number of recognized species present in most ecosystems is also among the highest for any specific geographic locality representative of any bat family. Most of the vesper bats are insects feeders, only a few feed on fish and other aquatic invertebrates and they echolocate through their mouth (Nowack 1994). Interestingly, most (23 of 32) of the migratory species of the bats belong to the family Vespertilionidae (Bisson *et al.* 2009), suggesting an increased frequency of vesper bats, in particular of exposure to diverse environments. It is possible that some of these behaviors in conjunction with their long life and reproductive span (Neuweiler 2000, Vonhof *et al.* 2006) may be contributing factors. Bats are also exposed to various insect, plant and mammalian viruses (Li *et al.* 2010) and are frequently a reservoir species for large numbers of viruses (Calisher *et al.* 2006). Indeed, host-parasites interactions have been implicated in the horizontal introduction of DNA to genomes (Gilbert *et al.* 2010, Hotopp *et al.* 2007, Hecht *et al.* 2010). Vesper bats appear to possess some unknown combination of attributes that increase their susceptibility to the horizontal invasion of TEs, these attributes may also play a role in predisposing them as viral reservoirs.

2.4.3 Implications of *Helitron* colonization in the vesper lineage

The invasion and colonization of *Helitrons* in vesper genomes (≈3%) have had a tremendous impact on genome diversification. Because of their ability to capture and amplify gene fragments, *Helitrons* may have played an incredible role in altering the genomic landscape. Indeed, *Helitrons* have amplified the promoter and the first exon of the highly conserved single copy gene, nucleotide binding protein-like (NUBPL) to ≈1000 copies in the *M. lucifugus* genome (Pritham and Feschotte 2007). The estimated age of amplification of *Helitrons* and other DNA transposons in the vesper lineage roughly coincides with the diversification of vesper bats (16-25 mya; Pritham and Feschotte 2007, Oliver and Greene 2009, Zeh *et al.* 2009). In fact, the Vespertilionidae is the most speciose family among mammals, (>400 species), second to Muridae (Old world rats and mice), (Simmons 2005).

Abundant diversification has also occurred in the family Phyllostomidae, which display the greatest morphological diversity within a bat family. Phyllostomid bats have a tremendous

variability in morphology that is directly linked to feeding strategies. While, most bat families (14 of 19) are strict insectivores and insectivory is considered as the primitive condition for bats (Baker *et al.* in press), phyllostomid species have adapted to feeding as strict insectivores, omnivores, sanguivores, carnivores (eating other vertebrates), nectar feeders and fruit eaters (Baker et al in press). *Helitrons* were not detected in various phyllostomid samples we analyzed, which represented these diverse feeding habits. Morphologically, vespertilionids are variable around the theme of being an insectivore. Massive amplification of *Helitrons*, followed by colonization of diverse DNA transposons could have generated tremendous variation in genome architecture and landscape aiding in the diversification within the vesper lineage, perhaps enabling the adaptation to diverse environments.

## 2.5. Conclusions

To summarize, our study provides the evidence for the restricted invasion and amplification of the *Helibat* family of *Helitrons* in the vesper bat lineage. Indeed, we were not able to identify *Helitrons* in different bat species representing all four super groups belonging to two sub-orders of bats, by employing *in silico*, PCR and hybridization based techniques. The discontinuous distribution of *Helitrons* observed across the different bat families adds further support to the hypothesis that horizontal transfer is a major player in the introduction of *Helitrons* to the vesper bat lineage. The invasion and colonization of *Helitrons* with their ability to capture and amplify gene fragments could have tremendously impacted on the genome evolution and diversification of vesper bats.

CHAPTER 3

# IMPACT OF *HELITRON* AMPLIFICATION ON THE GENOME ARCHITECTURE OF THE MAMMAL, THE LITTLE BROWN BAT

## 3.1 Introduction

Transposable elements (TEs) constitute a significant portion of many eukaryotic genomes. They have the unique ability to move and replicate within the genome. Based on the transposition intermediate, TEs are classified as class 1 retrotransposons and class 2 DNA transposons (Craig 2002). Retrotransposons which utilizes RNA intermediate are further classified into Long Terminal Repeat (LTR) retrotransposons, non LTRs and DIRS. DNA transposons have DNA intermediate and are further divided to classic cut-and-paste DNA transposons, *Helitrons* and *Mavericks* (Pritham 2009, Feschotte and Pritham 2007). Because of their dynamic nature, TEs can engender dramatic changes to genome architecture in a myriad of ways. Among those, transduction and further amplification of genic fragments by TEs have received widespread attention recently, because of its implications genome evolution. TE mediated transduction and duplication can create novel genetic units and can alter the genome architecture (see review Feschotte and Pritham 2007). Nearly, all major groups of TEs have been reported to have transduced host genic sequences (Moran *et al.* 1999, Ejima and Yang 2003, Xing *et al.* 2006, Jiang *et al.* 2004, Lal *et al.* 2003, Lai *et al.* 2005). However, *Helitrons* are the most notorious among TEs and have captured and shuffled genome sequences at a remarkable rate (see review Feschotte and Pritham 2007).

The remarkable ability of *Helitrons* to capture and amplify gene fragments has been occasionally attributed to their mechanism of transposition (Feschotte and Pritham 2007). *Helitrons* encode a putative protein with a rolling circle initiator motif and PIF1-like DNA helicase

domains (Kapitonov and Jurka 2001). Homology of the proteins encoded by *Helitrons* with the single-stranded bacteriophages, plant gemini viruses, and other bacterial plasmids which undergo rolling circle replication supports the hypothesis that this mechanism facilitates the transposition of *Helitrons* (Kapitonov and Jurka 2001, Feschotte and Wessler 2001). *Helitrons* are also unique from other DNA transposons in their structure in addition to the mechanism of transposition. *Helitrons* have a TC at the 5' terminus and a CTRR at the 3' termini. They lack terminal inverted repeats and do not produce any target site duplication upon transposition typically insert between an AT dinucleotide.

*Helitrons* constitute a significant portion of many eukaryotic genomes. They contribute 2-3% of the *Caenorhabiditis elegans*, *Arabidopsis thaliana* (Kapitonov and Jurka 2001) *Zea mays* genomes (Yang and Bennetzen 2009). A single of family of *Helitron* known as *Helibats* harbors 3-4% of the *M. lucifugus* genome, the only mammal that harbours *Helitrons* (Pritham and Feschotte 2007) and 5% of the *Drosophila virilis* (Kapitonov and Jurka 2007) genome, the highest contribution by *Helitron* reported to date. *Helitrons* have a disparate distribution across the eukaryotic tree of life. However, capture of gene fragments has been reported in many organisms (Cultrone *et al.* 2007, Hollister and Gaut 2007, Sweredoski *et al.* 2008, Choi *et al.* 2007, Choi *et al.* 2007, Langdon *et al.* 2009, Pritham and Feschotte 2007, Yang and Bennetzen 2009a) even though, this ability is particularly pronounced in maize (Sweredoski *et al.* 2008). *Helitrons* have shuffled >20,000 gene fragments in the maize genome and have created novel genetic units (Morgante *et al.* 2005 Du *et al.* 2009, Feschotte and Pritham 2009, Yang and Bennetzen 2009b, Lal *et al.* 2003, Brunner *et al.* 2005, Gupta *et al.* 2005, Lal and Hannah 2005, Xu and Messing 2006). The activity of *Helitrons* in maize has also resulted in intra-species diversity and in the loss of genic co-linearity (Lai *et al.* 2005). The captured gene fragments are sometimes transcribed resulting chimeric or mosaic transcripts and could impact gene expression in several ways (see review Lal *et al.* 2009).

In *M. lucifugus, Helitron* has captured and amplified the promoter, 5' UTR and first exon of the NUBPL gene, a highly conserved single copy mammalian gene (Pritham and Feschotte 2007). *Helitron* along with the captured fragment has amplified to >1,000 copies, in the little brown bat genome. It is intriguing to speculate that gene capture is a general phenomenon of *Helitrons* that has contributed to the evolution of the genome architecture of this species. Even though there are numerous examples of gene capture by *Helitrons* in different organisms, the mechanism of gene capture remains unknown in part because we still do not fully understand how *Helitrons* replicate. Here, we provide a comprehensive analysis of the *Helitrons* in the genome of *M. lucifugus*, their diversity, abundance and the nature and pattern of the gene captures and propose a model for gene captures based on the pattern observed.

## 3.2 Methods

### 3.2.1. Identification and classification of *Helitrons*

#### 3.2.1.1 *de novo* repeat identification

The 7x version of *M. lucifugus* genome (GL429767-GL433173, 3407 scaffolds, Sanger sequencing) was downloaded from the WGS database at National Center for Biotechnology Information (NCBI) to a local server. The estimated chromosome number for the genus *Myotis* is 2n=44 (Ao *et al.* 2006) and estimated genome size is ~2.3Gb (Pritham and Feschotte 2007, T. R. Gregory, Animal Genome Size Database; www.genomesize.com). The sequence data summed upto 1.91 Gb coveres ~83% of the total *M. lucifugus* genome. The *de novo* repeat identification software Repeatscout 1.0.2 (Price *et al.* 2005) was employed to identify all the repeat families using default parameters, which require that the DNA occur in at least 3 copies that are at least 50 bp per genome. To further assemble repeat families sharing sequence identity ≥90% with a minimum overlap of 100 nucleotides, we utilized the assembly function of the software package Sequencher version 4.7 (Gene Code Corporation, Ann Arbor, Michigan).

3.2.1.2 *Helitron* discovery and classification

The assembled repeats were subjected to automated TE classification based on structure and homology using the software Repclass (Feschotte *et al.* 2009). To develop a preliminary library, repeats identified as *Helitrons* were validated by eye using BLAST tools (Altschul *et al.*1990) to eliminate false positives and delimit the proper boundaries. This library of *Helitrons* was then used to mask the assembled repeat families identified by Repeatscout. RepeatMasker, a program that can screen DNA based on homology (RepeatMasker version 3.2.7; A. F. A. Smit, R. Hubley, and P. Green, www.repeatmasker.org) was employed to find any additional related elements missed by the first screen. New repeats that were masked by *Helitrons* were in turn curated by eye and added to the library. Curated *Helitrons* were classified into families based on sequence identity in the last 30 bp (at least 80%), as subfamilies based on the sequence identity (at least 80%) at the first 30 bp and exemplars based on their divergence (>20%) at the internal region (Yang and Bennetzen 2009). Classification of *Helitrons* into exemplars was done using cd-hit-est which can cluster sequences based on nucelotide identity (Huang *et al.* 2010).

3.2.1.2 Empty site identification

To illustrate the mobility and to validate the structural boundaries of novel families, subfamilies and of *Helitrons* with gene fragments, empty sites were identified. A chimeric query was constructed after extracting 50 bp upstream and 50 bp downstream flanking the *Helitron*s. To determine the sequence state prior to the insertion of the element nucleotide based searches (blastn default parameters) were employed to identify regions homologous to the chimeric query within the *M. lucifugus* genome (Paralogous) or in closely related mammals (Orthologous) when possible.

3.2.2 Identification of host genomic sequences within *Helitrons*

*Helitrons* are well known to capture gene sequences, which in some cases can be further amplified in the genome. To identify cases of gene capture by *Helitrons,* homology based searches were employed. To determine regions of putative capture, the curated *Helitron*

library was used to query all mammalian genomes deposited in the whole genome shot gun sequences (WGS) and nucleotide collection (nr/nt) at NCBI excluding the *M. lucifugus* genome using nucleotide based searches (Blastn, default parameters). As these genomes are devoid of *Helitrons*, the only positive hits would be to regions homologous to non-*Helitron* regions conserved in other mammalian genomes. The resulting hits (e-value $\leq 10^{-04,>}$, >50 bp) were then used as queries against the nr/nt or USCS genome browser to determine if any of the captured regions were annotated in the human genome. Since *M. lucifugus* genome is not yet annotated, human gene information from the UCSC genome browser was used for characterizing the captured region as potential promoter, untranslated region (UTRs), coding exons or introns. One of the possible limitations with our method is that genes that are specific to *M. lucifugus* or that are subjected to high mutational changes cannot be detected. However, our methods are robust enough to identify capture of the genomic sequences that are conserved at least in closely related mammals, considering the slow rate of sequence decay in mammals (Yi *et al.* 2002, Teeling *et al.* 2005)

3.2.2.1 Identification of potential mechanism of gene acquisition

To understand the possible mechanism of the capture, the parental copy of the captured gene was identified when possible. To this end, an artificial query of the captured fragment and its nearby conserved region which was not captured was subjected to reciprocal blastn against the *M. lucifugus* genome. The contig or scaffold containing both the captured region and the conserved region was considered as the parental copy for that captured region. The regions corresponding to 10kb upstream and downstream to the captured region in the parental copy were extracted to determine the relationship of any *Helitrons* pieces that were located close by and to aid in developing a model for how the capture occurred.

3.2.3 Estimation of copy number and abundance of *Helitrons* in the genome

To estimate *Helitron* copy number, the first and last 30 bp of all *Helitrons* were queried to the *M. lucifugus* genome using WUBLAST (Repeatmasker) and hits more >80% identical

were counted. Regions homologous to non-*Helitron* sequences including secondary insertions and captured TE sequences were removed to generate a library of unique *Helitron* sequences. To identify the TE sequences, the *Helitron* library was masked (RepeatMasker) by DNA transposons and retrotransposons in the *M. lucifugus* genome (Repbase http://www.girinst.org/repbase/). The library was curated by removing sequences which were >100 bp and >80% identical to the query TE and the captured host genomic sequences were not removed. The curated library was used to RepeatMask the genome to estimate the proportion of the genome occupied by *Helitrons*.

3.2.4 Simulation analysis

Simulation analysis was performed to find whether there any bias in identifying the 5' regions (upstream/promoter and UTR), exons and 3' UTR of a gene was exhibited when using our homology-based method and to investigate on the insertional preference for 5' region of gene. The 5' region of the gene including promoter (1000 bp upstream) and UTR, coding exons and 3' UTR sequences of a random set of 100 genes were downloaded from the Table browser in UCSC genome browser and were subjected to blastn analysis against the *M. lucifugus* genome. The percent identity and bp aligned were noted. In order to determine insertion preferences of *Helitron*, 500 bp upstream and downstream flanking *Helitron* insertions were extracted and blasted to the human genome. The target sequence along with flanking 1kb upstream and downstream was taken and subject to homology searches against the human genome assembly at the UCSC genome browser to further characterize the site of integration.

3.2.5 Tests of purifying selection

*Helitron* copies containing full length retrogenes were extracted and analyzed for intact open reading frame. Those with intact open reading frame were aligned with the parental copy when possible to identify whether *Helitrons* amplified copies are evolving under selection. If the parental copy was not identified, a consensus sequence was used as a proxy for the parental (consensus was constructed from all *Helitron* amplified retrogenes based on the majority rule)

A test for purifying selection was performed with the software package, MEGA (4.0.2*,* Tamura *et al.* 2007) using pairwise deletion and the Nei-Gojobori Method.

## 3.3 Results

### 3.3.1 Identification and classification of *Helitrons*

In order to identify all the families of *Helitrons* in the *M. lucifugus* genome sequence, a combination of both homology and *de novo* based approaches were employed. As a starting point, all repeat families (≥ 3copies) were identified in the 7X coverage of the *M. lucifugus* genome (Accession AAPE00000000) using the program Repeatscout (Price *et al.* 2005). In total, 30,655 repeat families were identified, which were further assembled into 15859 repeats using the assembly program, sequencher. To automate the classification of these repeats using both homology and structural features, the program REPCLASS was employed (Feschotte *et al.* 2009). REPCLASS classified 176 repeats as *Helitrons*. Each of the putative *Helitron*s was in turn, were validated by eye using homology based methods. Using this set of techniques, a library of 105 *Helitrons* was constructed. To identify other repeats in the assembled repeat library related to the curated *Helitrons*, the assembled repeat library was masked with the *Helitron* library. Using this approach another 3231 repeats were tentatively identified as *Helitrons*. In total, a library consisting of 477 intact *Helitrons* was generated. *Helitrons* were classified into 38 families (Table 1) based on their identity (at least 80%) over the last 30 bp and 59 subfamilies based on their identity (at least 80%) over the first 30 bp (Yang and Bennetzen 2009). Clustering (Huang *et al.* 2010) allowed 477 *Helitrons* to be grouped into 415 exemplars based on the divergence at the internal region (>20%). Orthologous or paralogous empty sites were identified for 31 families, 24 subfamilies to validate boundaries of the novel families (Table 3.1, Appendix F). Since we analysed repeats that are more than three copies, we could have missed single copy *Helitrons* in the *M. lucifugus* genome.

Table 3.1. Characteristics of *M. lucifugus Helitron* families

| Family name | copy number | Putative autonomous | Gene fragments | Empty site confirmation |
|---|---|---|---|---|
| *Helibats* | 153334 | (+) | (+) | (+) |
| *HelibatN9* | 24 | (-) | (-) | (+) |
| *HelibatN8.1a* | 3 | (-) | (-) | (+) |
| *HeligloriaB* | 21 | (-) | (-) | (+) |
| *HelibatN_10t* | 10 | (-) | (-) | (+) |
| *HelibatN1.2b* | 283 | (-) | (+) | (+) |
| *HelibatN6.2* | 1705 | (-) | (-) | (+) |
| *53* | 2 | (-) | (+) | (+) |
| *63* | 27 | (-) | (-) | ND* |
| *78* | 1 | (-) | (-) | (+) |
| *215_a* | 8 | (-) | (-) | (+) |
| *103* | 1 | (-) | (-) | ND* |
| *129* | 2 | (-) | (-) | ND |
| *234* | 2 | (-) | (-) | (+) |
| *235* | 51 | (-) | (-) | (+) |
| *133* | 101 | (-) | (-) | (+) |
| *149* | 5 | (-) | ND | ND* |
| *241* | 5 | (-) | (-) | (+) |
| *153* | 6 | (-) | (-) | (+) |
| *154* | 41 | (-) | (-) | (+) |
| *160* | 2 | (-) | (-) | (+) |
| *245* | 1 | (-) | (-) | (+) |
| *172_n* | 5 | (-) | (-) | ND* |
| *247* | 497 | (-) | (-) | ND |
| *250* | 1 | (-) | ND | (+) |
| *184* | 8 | (-) | ND | (+) |
| *190* | 21 | (-) | ND | (+) |
| *191* | 89 | (-) | (-) | (+) |
| *192* | 6 | (-) | ND | (+) |
| *198* | 1 | (-) | (-) | (+) |
| *200* | 4 | (-) | (-) | (+) |
| *HelibatN3.2* | 1 | (-) | (+) | (+) |
| *HelibatNA27* | 12 | (-) | (-) | (+) |
| *HelibatN1.5c* | 5 | (-) | (-) | (+) |
| *236* | 1 | (-) | (-) | (+) |
| *HelibatTT* | 1 | (-) | (+) | ND |
| *HelibatN5.1b* | 469 | (-) | (-) | (+) |
| *238* | 7 | (-) | (-) | (+) |
| *HHsearch12* | 4 | (-) | (-) | ND |

* Ends of this element is flanked by another *Helitron*

3.3.2 Structure, copy number, and abundance of *Helitrons*

The size of the identified *Helitrons* varied from 151-5503 bp. Most of the identified *Helitrons* (91%) belonged to the *Helibat* family, which is the only family that has a putative autonomous partner (*Helibat1*; Pritham and Feschotte 2007). None of the 37 novel identified familes have a putative autonomous partner (see Table 1). It is possible that these elements could have transposed using *Helibat* machinery, however there are no evidences available suggesting that. Copy number and abundance of *Helitrons* were estimated by employing WUBLAST and RepeatMasker respectively. The number of 5' ends and 3' ends in the *M. lucifugus* genome was 121,722 and 156, 462 respectively. Some *Helitrons* contain TEs sequences (secondary insertions or capture), since their repetitive nature could influence the estimation of amount of DNA contributed to the *M. lucifugus* genome, TEs (>100 bp and >80% identical to the *M. lucifugus* TE library, Repbase) within the *Helitrons* were removed. However, the gene fragments within the *Helitrons* were not removed as their influence on the estimation could be negligible due to their low copy number. Masking with the curated *Helitron* library revealed that it contributed to 219.3 Mb of *M. lucifugus* genome (~11.5% of sequenced region ~1.9GB, estimated genome size ~2.3 GB). This represents the largest amount of genomic DNA contributed (~4 fold, ~45.5 Mb in maize; Yang and Bennetzen 2009) generated by *Helitrons* ever reported. Their estimate was done after removing DNA transposons and retrotransposons identified in maize from *Helitrons*.

3.3.3 Capture of host genomic sequences

To identify *Helitrons* that have captured and further amplified gene fragments, blastn searches were performed using the entire library of *Helitrons* against the WGS and nr/nt databases excluding the *M. lucifugus* genome. *Helitrons* with significant hits (≥50 bp with an e-value $\leq 10^{-04}$) were subjected stringent manual analyis for the presence of host genomic sequences. Captured genomic sequences were classified as genic, non-genic and TEs. Regions that were classified as genic were homologous to promoters, UTRs, introns or exons of

known human genes. Non-genic region refers to captured regions that could not be readily allied to any known gene or conserved regions, but could be identified as part of the host genome. Through this technique, 24 cases of genic captures and eight cases of non-genic captures were identified. *Helibat* family was linked to 28 cases of gene captures and four novel families were associated with four cases of gene captures (see Table1). However, this is likely to be still an underestimate of captured DNA sequences since detection methods relies on homology with with other genome sequences. Our detection methods fail to identify *M. lucifugus* specific sequences and genes that are subjected to high mutation rate. Paralogous or orthologous empty sites were identified for 26 of the 32 gene captures identified (Appendix F). In addition, three *Helitrons* contained a portion of another TE that appears to have been captured. These examples do not show the hallmarks of nested transposition events such as the occurrence of target site duplications (TSDs) flanking complete TE insertions.

3.3.4 Characterization of the gene captures

3.3.4.1 Structure and nature of gene captures

Each *Helitron* family identified was examined closely to determine both the structure and boundaries of the sequences that were captured. To this end, we sought to identify the region of the *M. lucifugus* genome from where the sequence was captured (parental region) as well as the region orthologous to the parental region in the human genome. The human genome is the best annotated mammalian genome and provided an excellent source of information as to the structure of homologous genes. The slow mutation rate allows a comparison to be made between even between distantly related mammals. The sizes of the captured fragments detected including both genic and non-genic regions varied from 86 bp to 1636 bp, a range consistent with size of *Helitrons* identified in the genome (151-5503 bp). In total, we identified 32 capture events involving five *Helitron* families. *Helitrons* have captured and amplified 5' region (promoters [upstream 1000 bp], 5' UTR, first exon and intron) of 13 genes, 3' region (3' UTR) of three genes and internal region (introns or exons) of five genes. Since there is no

information available on the gene structure in *M. lucifugus*, these regions are predicted soley based on homology with the human genes. However, It is possible that the structure of gene could be could be different in *M. lucifugus* genome. Interestingly, our results revealed an apparent bias in the region of genes captured. The most frequent regions captured were promoter (upstream region 1000bp), 5' UTR, and first exon. The apparent bias in the captured region could due to a bias in identifying the 5' region of the genes better than the 3' region of the genes or could be due to an insertional preference of *Helitron* towards the 5' region of the gene. To determine whether this pattern was due to a bias in identifying the 5' region of the gene compared to the 3' region, a simulation analysis was performed by employing blastn searches of the 5' UTR and 3' UTR of 100 random human genes against *M. lucifugus* genome. These analyses revealed no bias for identifying either the 5' region or 3' region of genes. To determine whether the observed bias for 5' region was caused by a general preference for *Helitron* insertion, 145 random insertions were analyzed. Since *M. lucifugus* genome is not annotated, flanking regions were compared with the human genome. 37% of the insertions analyzed where in the intronic regions of the genes and 31% within TEs and 0.6% was in the 5' UTR and 3' UTR of the genes.

Table 3.2. Characteristics of gene fragments acquired by *Helitrons*

| Name of *Helitron* (Length in bp) | Captured fragment | Region captured | Orientation | Size (bp) | % ID | Closely related organism | Copy # |
|---|---|---|---|---|---|---|---|
| *HelibatN2* (2174) | 5,10-methenyltetrahydrofolate synthetase (MTHFS), | 5'UTR, exon,intron | + | 276 | 70 | *Pteropus vampyrus* | 471 |
| *HelibatN_Stk24_mor4* (1859) | Serine threoinie kinase 24 isoform b | Promoter, exons, 5'UTR, intron | - | 1369 | 72.5 | *P. vampyrus* | 104 |
| *HelibatN1.26* (686) | E1A binding protein p400 (EP400)<br>DENN/MADD domain containing 5B (DENND5B), | Intron, Exon<br>5'UTR and 1st Exon | +<br>+ | 206<br>86 | 75<br>77 | *Felis catus*<br>*Homo sapiens* | 72 |
| *HelibatN1.DD* (690) | DENN/MADD domain containing 5B (DENND5B), | promoter,5'UTR and 1st Exon,intron | + | 540 | 70 | *Tursiops truncatus* | 255 |
| *HelibatN1.21_a* (711) | E1A binding protein p400 (EP400) | Exon | + | 106 | 91 | *Homo sapiens* | 50 |
| *HelibatN1.2b* (618) | membrane-associated ring finger (C3HC4) 5 (MARCH) | 5'UTR and 1st Exon | + | 301 | 75 | *P. vampyrus* | 102 |
| *Stat_3* (1891) | protein inhibitor of activated STAT, 1, transcript variant 2 (PIAS1) | Promoter, 5'UTR, exon, Intron | - | 1238 | 71 | *Pan troglodytes* | 194 |
| *>53* (861) | GNAS complex locus | Intron,Exon | + | 597 | 76 | *Equus caballus* | 104 |
| *Helibat1.5q_N1* (2321) | Kv channel interacting protein 1 (KCNIP1) | Intron | - | 1521 | 69 | *E. caballus* | 126 |

Table 3.2 continued

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *HelibatN1.5n* (2503) | proteasome activator subunit 3 isoform 1(PSME3)<br>TTBK2 | 3'UTR<br>Intron | -<br>- | 608<br>282 | 75<br>77 | *P.troglodytes*<br>*T. truncatus* | 173 |
| *HelibatN1.5a3_bat1* (1430) | microfibrillar-associated protein 1 (MFAP-1) | 5'UTR, Exon, intron | + | 968 | 75 | *P. vampyrus* | 230 |
| *HelibatN1.5t_N2* (1140) | Tau-tubulin kinase 2 (TTBK2) | intron | - | 446 | 71 | *T. truncatus* | 118 |
| *>HelibatTMBIM4* (2489) | transmembrane BAX inhibitor motif containing 4 (TMBIM4) | 5'UTR, exon, intron | - | 235 | 69 | *H. sapiens* | 13 |
| *HelibatCCB1* (2462) | transforming, acidic coiled-coil containing protein 3 (TACC3)<br>transmembrane BAX inhibitor motif containing 4 (TMBIM4) | 4 exons , 3' UTR<br>5' UTR,exon, intron, | +<br>- | 327<br>241 | 74<br>74 | *Equus caballus, Microcebus murinus* | 118 |
| *46_N2* ( 2519) | SFRS protein kinase 1 (SRPK1) | promoter and 5'UTRs | - | 1262 | 73 | *P. vampyrus* | 1 |
| *HelibatCCB2_only_TACC3* (2043) | transforming, acidic coiled-coil containing protein 3 | Exons, 3'UTR | + | 454 | 71 | *Canis familiaris* | 26 |
| *HelibatTT* (556) | Transducin beta-like 1X-related protein 1,TBL1XR1 | Promoter | - | 272 | 82 | *Gorilla gorilla* | 90 |
| *HelibatN4.2* (911) | sphingomyelinase (N-SMase) activation associated factor(NSMAF) | 5'UTR | - | 134 | 75 | *Mus musculus* | 21 |
| *HelibatN2.5b_tandem*(1830) | mesoderm induction early response 1, family member 3 (MIER3) | 3'UTR | - | 199 | 83 | *Ailuropoda melanoleuca* | 26 |

Table 3.2 continued

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *174* ( 1947) | nephrocystin 1 isoform 3_NPHP1 | Promoter and intron | + | 1264 | 69 | *A. melanoleuca* | 67 |
| *HelibatN2.12b* (1275) | selenoprotein S SELS | Exon 1st | - | 130 | 75 | *A. melanoleuca* | 36 |
| *HelibatN2.15* (1894) | nucleolar protein with MIF4G domain 1 | exon and intron | - | 660 | 75 | *T. truncatus* | 550 |
| *HelibatN1.4a_TAPT1GC* (510) | Transmembrane anterior posterior transformation protein 1_TAPT1 | 5'UTR | + | 71 | 81 | *H. sapiens* | 241 |
| *HelibatN1.30_N2* (2582) | G patch domain containing 2 (GPATCH2) | intron | - | 1636 | 71 | *P. troglodytes* | 74 |
| *HelibatN1.17.1* (1958) | Unknown, conserved across placental mammals | NA | NA | 860 | 79 | *P. vampyrus* | 1045 |
| *HelibatNa10* (587) | Unknown conserved | NA | NA | 446 | 72 | *P. vampyrus* | 115 |
| *HelibatN1.3e* (1032) | unknown | NA | NA | 202 | 74 | *T. truncatus* | 81 |
| *HelibatN1.2a_a_N3* (659) | unknown | NA | NA | 394 | 75 | *P.vampyrus* | 1149 |
| *HelibatN1.30_z* (1053) | unknown | NA | NA | 536 | 68 | *Felis catus* | 9 |
| *194a* (1184) | unknown | NA | NA | 319 | 72 | *P.vampyrus* | 49 |
| *HelibatN2.11_a* (1459) | unknown | NA | NA | 312 | 71 | *T. truncatus* | 1105 |
| *214_GC* (558) | unknown | NA | NA | 318 | 74 | *E. caballus* | |

NA- not applicable

3.3.4.2 Examples of gene captures

3.3.4.2.1 PIAS1 capture by *Helitron*

Alignments of *Helibat_Hpstat* revealed significant sequence identity (71% over 1238 bp) with the protein inhibitor of activated stat1 (PIAS1) gene from chimp and human (Figure 1). This is strong evidence that a ~1238 bp region including the promoter (500 bp), 5' UTR (93 bp), first exon (24 bp) and part of the first intron (581 bp) in the reverse orientation was captured by a *Helibat* family of *Helitron*. There are 194 copies displaying >75% sequence identity over 50% of the captured fragment in the *M. lucifugus* genome. The presence of promoter and intron in the captured fragments indicates that the capture most likely occurred at the DNA level. The captured regions constituted ~63% of the total element. The PIAS1 inhibits DNA binding activity of Stat1 (signal transducer and activator of transcription 1) mediated gene activation. Stat1 is involved in the activation of specific genes including interferons, which are involved in the innate immunity (Liu *et al.*1998). PIAS1 gene is also a component of DNA damage response pathway (Galanty *et al.* 2009) and has a role in germ line development (Jones *et al.* 2006).

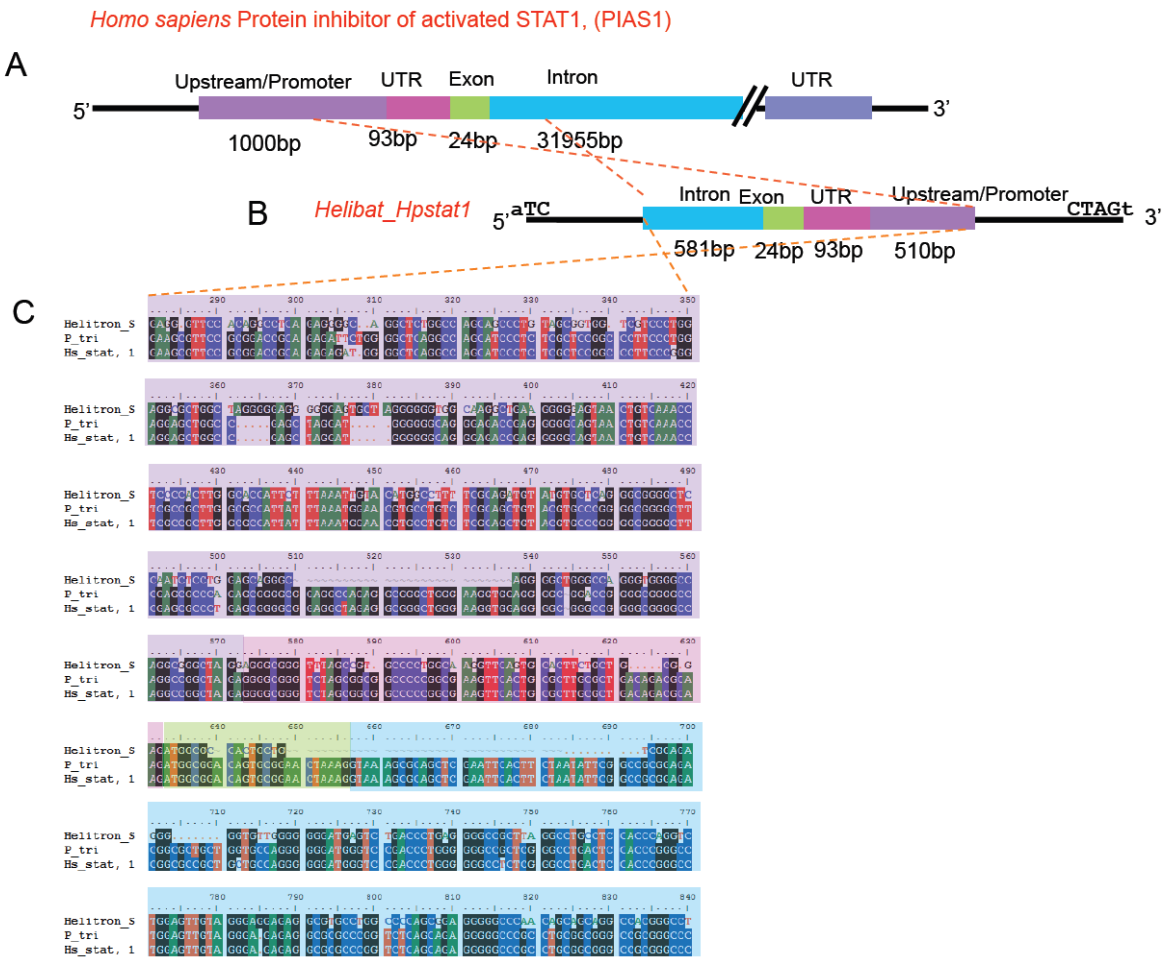



Figure 3.1. A schematic representation of *Helibat_Hpstat1,* which contains a fragment of the protein inhibitor of activated STAT1 gene. A) The structure of the human PIAS1 gene. B) The structure of *Helibat_Hpstat1*, the *Helitron* containing the gene fragment of PIAS1 and C) An alignment of part of the promoter, 5' UTR, exon and intron of the gene in human and chimp to the gene fragment in *Helibat_Hpstat1*.

3.3.4.2.2 Stk-24 capture by *Helitron*

Alignment of the *Helitron, HelibatN_stk24* with the Stk24 gene of Human and mega bat, *Pteropus vampyrus* revealed 72.5% identity over 1369 bp. *HelibatN_stk24* (*Helibat* family) has captured the promoter (827 bp), 5'UTR (400 bp), first exon (42 bp) and part of the first intron (79 bp) of the stk 24 in the reverse orientation (Figure 3.2). This gene fragment was amplified to 187 copies in the genome displaying >75% identity over 568 bp of the captured fragment. The captured region constituted ~73% of the *Helitron* and is in the reverse orientation. In human the encoded Stk24 protein is involved in protein phosphorylation, signal transduction and has brain specific expression (Zhou *et al.*, 2000).

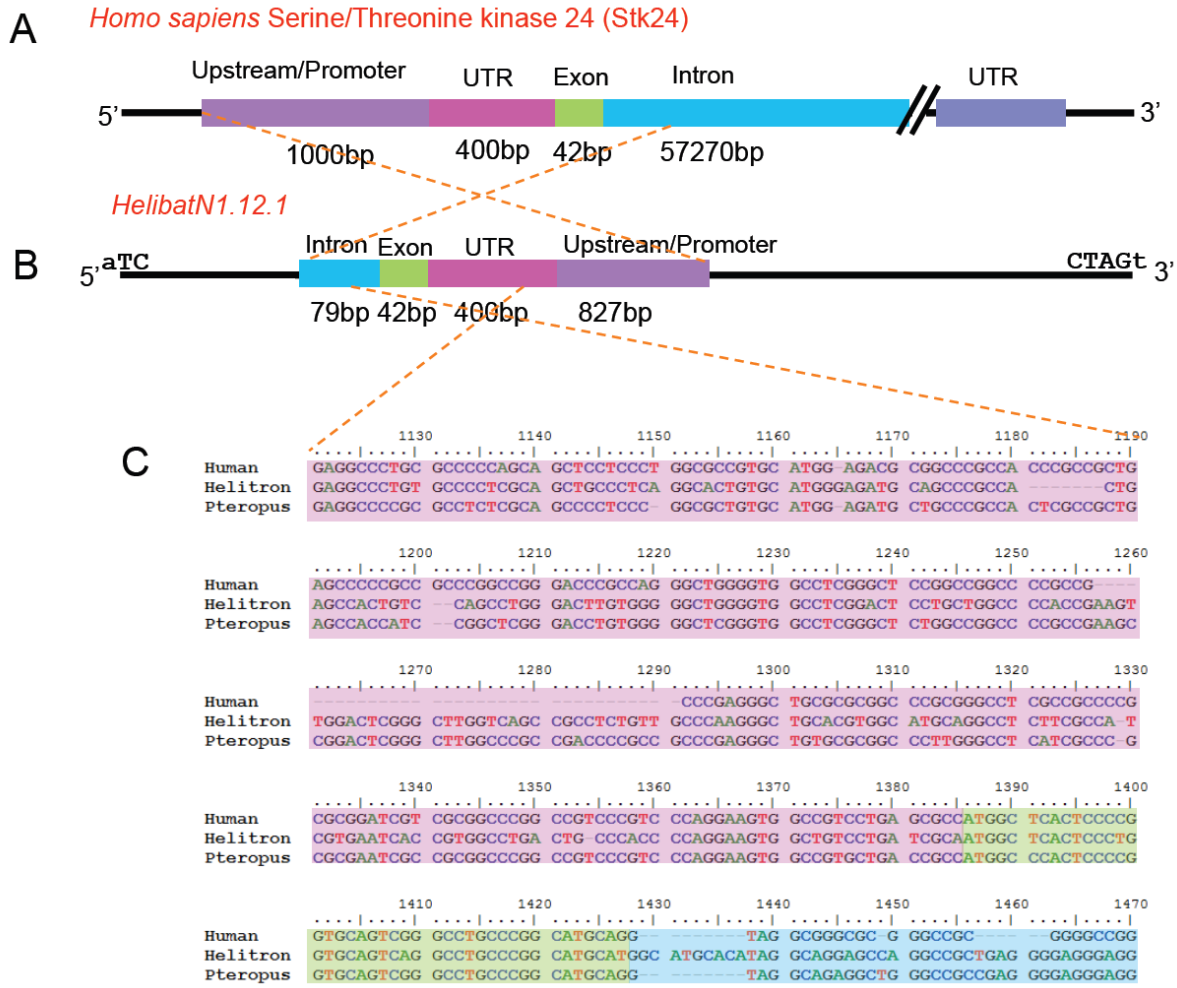



Figure 3.2. A schematic representation of *HelibatN1.12.1*, the *Helitron* containing a fragment of Stk24. A) The structure of the human Stk24 gene. B) The structure of the *HelibatN1.12.1*. C) An alignment of the portion of the 5' UTR, exon and intron of the Stk24 gene from human and mega bat to the gene fragment captured by *HelibatN1.12.1*.

3.3.4.3 Capture of multiple gene fragments

3.3.4.3.1 Capture of TACC3 and TMBIM4 gene fragments

*Helitrons* also contain regions from multiple loci and three such cases were identified. Fragments of both the transforming acidic coiled-coil containing protein 3 (TACC3) retrogene and transmembrane BAX inhibitor motif containing 4 (TMBIM4) gene (Table 3.2) were identified in *HelibatCCB1* (Figure 3.3). An alignment of *HelibatCCB1* with TACC3 mRNA in human and the TACC3 retrogenes identified from *M. lucifugus* and *Equus caballus* revealed significant sequence identity (74% identity over 327 bp). *Helibat CCB1* contains four exons and the 3' UTR of TACC3 gene in plus orientation (Figure 3.3) and has amplified to ~16 copies displaying >90% identity over 460 bp. TACC3 gene is shown to have some role in stabilizing spindle microbtubules (Gergely *et al.* 2003) and is expressed in the germ line (Hao *et al.* 2002). Alignment of *HelibatCCB1* also with TMBIM4 gene in human and gray mouse lemur, *Microcebus murinus* unveiled 74% identity over 241 bp. This is a strong evidence that suggests *HelibatCCB1* has captured 241 bp including 5' UTR, first exon and a part of the intron of TMIBIM4 gene in the reverse orientation (Figure 3.3). TMBIM4 fragment have been amplified to >100 copies displaying >75% identity over 50% of the size of captured fragment (Table 3.2). TMBIM4 is predicted as a membrane protein that is involved in protein binding (GO: 0005515). *Helitron* named *HelbatCCB2_only TACC3* belongs to *Helibat* family contains only fragments of TACC3 gene and *HelibatTMBIM4* which is also a member of the *Helibat* family contains only TMBIM4 gene.

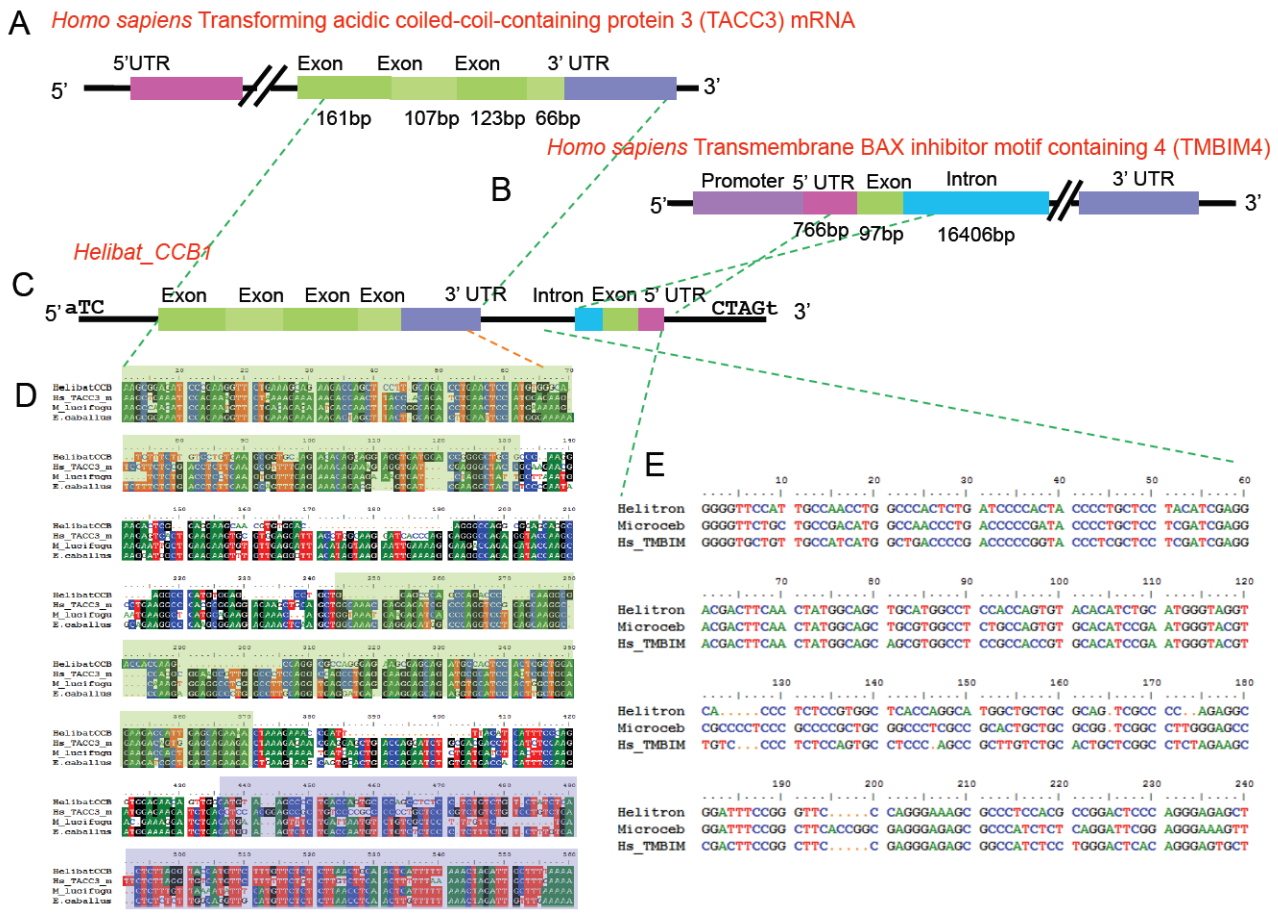



Figure 3.3. A schematic representation of a *Helitron* containing multiple gene fragments. A) The structure of the Human TACC3 mRNA. B) The structure of the Human TMBIM4 gene. C) The structure of the *Helibat_CCB1*, the *Helitron* that contains the TACC3 and TMBIM4 gene fragments. D) An alignment of the part of the TACC3 mRNA in human, TACC3 retrogene in *M. lucifugus* and *E. caballus* with the respective gene fragments. e) An alignment of the part of the UTR, exon and intron TMBIM4 gene in human, *Microcebus murinus* and the fragment in *Helibat_CCB1.*

3.3.4.3.2 Capture of EP400 and DENND5B gene fragments

The same pattern of presence of both multiple and individual gene fragments within *Helitrons* can be observed with capture of E1A binding protein (EP400) and DENN/MADD domain containing 5B (DENND5B). *Helibat N1.26* has captured an internal exon of EP400 gene and alignment with that of *Felis catus* revealed 75% over 206 bp. *HelibatN1.26* has also captured the 5' UTR and first exon of DENND5B gene and alignment with human gene revealed 77% identity over 86 bp. The *Helibat N1.26* has been amplified to ~300 copies displaying more than 77% identity over 50% of the *Helitron* in the *M. lucifugus* genome. DENND5B gene, which is predicted to function as a transmembrane protein (GO: 2444273). *HelibatN1.DD*, a *Helibat* family member captured only DENND5B gene and *HelibatN1.21_a* is also a *Helibat* family member has captured only EP400 gene which is involved in the transcriptional activation of genes (Table 3.2).

3.3.4.3.3 Capture PSME3 and TTBK2 gene fragments

*Helibat N1.5n,* has captured a fragment of the proteasome activator subunit 3 (PSME3) gene and Tau tubulin Kinase 2 (TTBK2) gene. Alignment of *HelibatN1.5n* with PSME3 gene of chimp reveals 75% identity over 608 bp and captured fragment (3'UTR) is in the reverse orientation. PSME3 encodes the structural component of immunoproteasome (Barton *et al.* 2004). Alignment of *HelibatN1.5n* with the captured intron of the TTBK2 gene in bottle nose dolphin reveals 77% identity over 282 bp. TTBK2 putatively phosphorylates tau and tubulin proteins and mutations in these gene can cause a neurodegenerative disease, spinocerebellar ataxia type 11 (SCA11; Houlden *et al.* 2007). Both genes were amplified to >100copies displaying >85% identity over 50% of the gene fragment. *HelibatN1.5t_N2* (*Helibat family*) contains fragments of only TTBK2 gene suggesting *Helitron* containing TTBK2 captured the PSME3 gene and *Helitron* copies containing only PSME3 gene were not identified in the *M. lucifugus* genome.

3.3.4.4 Identification of potential mechanism of gene acquisition

Lack of biochemical demonstration for the mechanism of transposition of *Helitrons* keeps the mechanisms of capture of host genomic sequences elusive. Howerver, there are some models proposed to explain the mechanism of gene captures. Feschotte and Wessler 2001 and Tempel *et al.* 2007 suggested proximity of *Helitrons* to the captured region as an essential criteria and proposed that capture occurred either through a inefficient recognition of the termination signal or by chimeric transposition of two different *Helitrons,* capturing the host genomic sequences in between. To identify presence of *Helitrons* near parental copy, 10kb upstream and downstream of the captured region was extracted and visually inspected. However, *Helitron*s were spotted only in the flanking region (within 10kb) of TMBIM4 parental gene among 13 cases of parental copies identified. *Helitron* containing TMBIM4 gene capture (*Helibat_TMBIM4.2*) is a chimeric *Helitron*, containing a short a non-autonomous full-length *Helitron* at the 3' termini. In the parental copy of TMBIM4 gene, downstream of the region captured by the *Helitron*, presence of a full length *Helitron* (progenitor) was identified (Figure 3.4). *Helitrons* copies that have captured TMBIM4 gene fragment (*Helibat_TMBIM4.2*, *Helibat_TMBIM4.1*) have the same 3' ends as that of the progenitor *Helitron* (Figure 3.4). The 5' end of the *Helibat_TMBIM4.2* and *Helibat_TMBIM4.1* is similar to *Helitron* copy located ~7000 bp upstream from the captured fragment in the parental copy. Different DNA transposons and retrotransposons were identified within 7000bp region but no regions of exons could be identified within that region suggesting the recent insertion of TEs in the respective region.

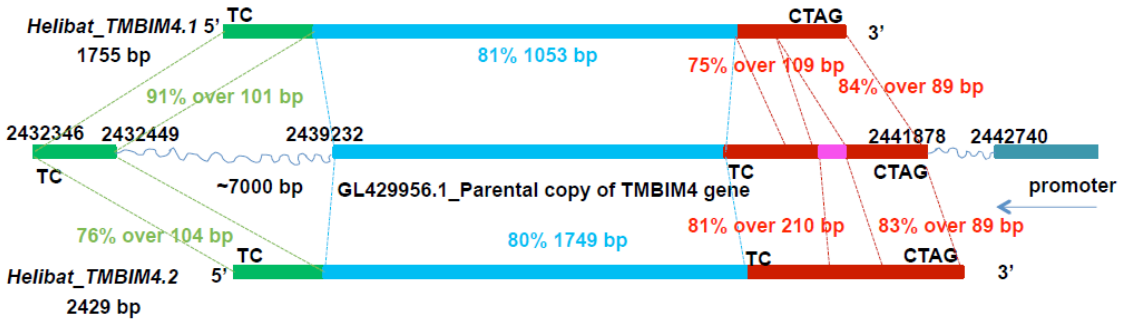



Figure 3.4 The analysis of the *Helitron*-captured region of the TMBIM4 gene in *M. lucifugus*. *Helitrons* containing the TMBIM4 gene fragment (Helitron_33_N1 and Helibat_TMBIM4) are compared with the parental gene TMBIM4

3.3.5 *Helitron*-mediated amplification of retroposed genes

The generation of retroposed gene copies is a common and well described phenomenon involved in generating gene duplicates in mammalian genomes (see review Zhang 2003). Our analysis uncovered five members of *Helibat* family (*HelibatN1.3c, HelibatN1.24_N2, Helibat_Ribo, HelibatNT_Ret and 40_N1*) in which retroposed mRNA transcripts could be readily identified. For each case, the signatures of retroposition including the lack of introns, a poly A tail and target site duplication (TSD) were systematically sought. *Helitrons* carrying these retroposed genes were further propagated (5-85 copies). In all five cases, the retrogenes were in the same orientation as that of the *Helitrons*. In three cases (*40_N1*, *Helibat_Ribo, HelibatN1.24_N2*) the regions amplified include 5' and 3' UTRs and coding exons where as in the other two cases (*HelibatN1.3c*, *HelibatNT_Ret*) only a part of coding exons and the 3' UTR. TSDs were identified for the retroposed gene, protein phosphatase 1 regulatory inhibitor subunit 12C (Ppp1r2c) containing coding exons and 3' UTR (Fig 3.5). This gene is a myosin-binding subunit of Protein phosphatase 1 (PP1) family which regulates PP1 activity and actin stress fiber assembly in a phosphorylation-dependent manner (Tan *et al.* 2001). *Helitron* has amplified Ppp1r2c retrogenes to 85 copies (displaying >80% over 50% of the retroposed gene fragment) in the *M. lucifugus* genome (see Table 3.3; Figure 3.5).

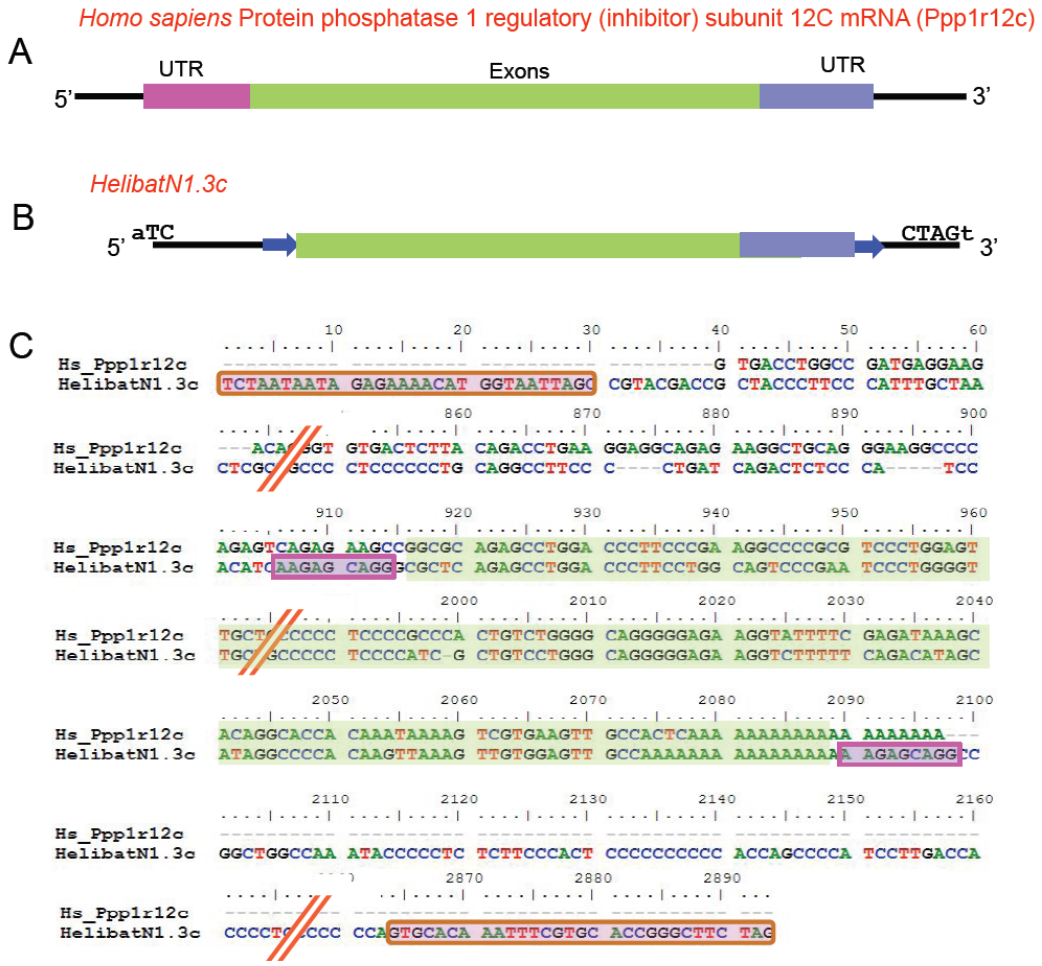



Figure 3.5. A schematic representation of a *HelibatN1.3c,* a *Helitron* carrying a retroposed copy of PPP1r12c gene. A. The structure of the human protein phosphatase 1, regulatory inhibitor subunit 12c mRNA. B. The structure of the *Helitron* with retroposed Ppp1r12c gene. C. An alignment of the *Helitron* containing the retroposed gene with the Human Ppp1r12c mRNA. TSDs are highlighted in pink boxes. The ends of the *Helitron* are highlighted in red boxes.

Table 3.3. Description of *Helitron* amplified retrogenes

| Name of the *Helitron* | Retroposed gene | Copy # | Regions | % ID | Size (bp) | Organism | Position within *Helitron* |
|---|---|---|---|---|---|---|---|
| *HelibatN1.3c* (2802) | Protein phosphatase 1, regulatory (inhibitor) subunit 12C (Ppp1r12c) | 85 | 7 Exons, 3' UTR, polyA,TSD | 80 | 1178 | *H. sapiens* | 872-2000 |
| *HelibatN1.24_N 2* (3140) | Nuclear prelamin A recognition factor isoform c, NARF | 50 | Part of 5'UTR, exons, 3' UTR | 78 | 1723 | *Canis famiiaris* | 584-2309 |
| *HelibatNT_Ret* (4804) | Pleckstrin homology domain containing, family G (with RhoGef domain) member 5 PLEKHG5 | 13 | 9 exons at 3' | 84 | 1884 | *H. sapiens* | 1089-2937 |
| *40_N1*(2197) | TCF3 (E2A) fusion partner (TFPT) | 45 | Part of 5'UTR, exons,3'UTR, polyA | 79 | 1126 | *H. sapiens* | 498-1592 |
| *Helibat_Ribo* (3402) | Ribosomal protein, large, P0 (RPLP0) | 5 | Part of 5'UTR, exons,3'UTR, polyA,TSD | 87 | 1100 | *H. sapiens* | 1310-2352 |

*HelibatNT_Ret* has retrogene fragments from multiple loci. *HelibatNT_Ret* contain part of a pleckstrin homology domain containing family G (PLEKHG5) gene and fragments of TACC3 retrogene and both of them are in the plus orientation. However, it cannot be confirmed whether the presence of TACC3 retrogene in the *Helitron* is due to a retroposition or capture event. Presence of TACC3 retrogene can be identified in the *M. lucifugus* genome which leaves the possibility of capture of the retrogene. PLEKHG5 gene encodes a protein that activates the nuclear factor kappa B (NFKB1) signaling pathway and mutations in this gene is found to be associated with distal spinal muscular atrophy (Maystadt *et al.* 2006, Maystadt, *et al.* 2007).

The nearly whole mRNA of nuclear prelamin A recognition factor isoform c, (NARF) ribosomal protein, large, P0 (RPLP0), and transcription factor 3 (TCF3) fusion partner, TFPT have retroposed to *HelibatN1.24_N2, Helibat_Ribo* and *40_N1* (*Helibat family*) respectively (Table 3.3). NARF is involved in the post-translational modification of proteins (Barton and Worman, 1999), RPLP0 encodes a ribosomal proteins that is a component of 60S subunit of ribosomes (RefSeq). TFPT, a highly conserved gene across mammals is involved in cell cycle progression, leukemogenesis and in programmed cell death induction (Brambillasca *et al.* 1999, Brambillasca *et al.* 2001, Irie *et al.* 2000, Gan *et al.* 2003, Franchini *et al.* 2006). *Helitron* has amplified the retroposed copy of the TFPT and NARF retrogenes to 45- 50 copies (displaying >85% sequence identity over 50% of retroposed gene fragment). RPLP0 is retroposed into a *Helitron* containing promoter, 5'UTR and a part of the first coding exon of the SFRS protein kinase (SRPK1) and *Helibat_Ribo* is amplified to five copies displaying >85% identity over >75% of the *Helitron.* Paralogous or orthologous empty sites were also identified for all five *Helitrons* with retroposed genes to confirm the ends of the element and validate the mobility of the element (Appendix F).

3.3.6 Retroposed genes evolving like pseudogenes

*Helitron* copies containing all full length retrogenes were tested for the presence of intact open reading frames. Two *Helitron* copies (40_N1) containing the TFPT retrogene were found to be intact and devoid of stop codons. Since the parental gene of TFPT could not be identified in the 7x version of *M. lucifugus* genome, a consensus of the gene from the 12 copies of *Helitron* (40_N1) containing the full length retroposed gene was constructed using a majority rule. The reconstructed copy was considered as the parental copy and was compared with the two *Helitron* copies with an intact TFPT gene. The dN/dS was measured between the two copies (0.64 and 0.98 p value >0.05) and the consensus and revealed that the two copies were evolving neutrally.

## 3.4 Discussion

3.4.1 *Helitrons* constitute a huge portion of the *M. lucifugus* genome.

This study provides the first comprehensive analysis of *Helitron* content in mammals. *Helitrons* are reported in a wide range of organisms and are known to constitute a significant portion (2-3%) of many genomes including both *A. thaliana* and *C. elegans* (Kapitonov and Jurka 2001). In maize, *Helitrons* constitute ~2.2% of the genome contributing almost 45.5 Mb DNA. In *Drosophila virilis* 180,000 copies of *Helitrons* were identified (Feschotte *et al.* 2009) estimated to make up ~5% (~9.4 Mb, 189.2Mb, 8X coverage) of the genome (Kapitonov and Jurka 2007). Our analysis identified ~156,000 *Helitron* copies in the *M. lucifugus* genome contributing ~219 Mb to the genome, the biggest contribution ever reported for *Helitrons*. This is remarkably different from other mammalian genomes sequences where there is no evidence of *Helitron* activity (Pritham and Feschotte 2007). It will be interesting to explore how *Helitrons* have altered the TE landscape of the *M. lucifugus* genome when compared to other mammals where DNA transposons constitute only a small fraction of the genome.

3.4.2 *Helitron* has captured and amplified >30 gene fragments

*Helitrons* have captured and amplified 32 gene fragments in the *M. lucifugus* genome, unveiling the propensity of animal *Helitrons* to capture gene fragments. The captured and amplified regions include promoter, UTRs, exons and intron of genes suggesting the role of *Helitron* amplified gene fragments in potentially altering the host-gene expression (see review Lal *et al.* 2009). By capturing and amplifying fragments from multiple gene fragments, *Helitrons* also shuffled promoters, UTRs and exons, playing a dynamic role in creating novel genetic units and potential regulatory subunits (Brunner *et al.* 2005, Gupta *et al.* 2005, Cultrone *et al.* 2007). Bat *Helitrons* tend to capture the 5' region of the genes more often compared to internal exons and introns where as in maize, internal exons are captured frequently (see review Kapitonov and Jurka 2007). This tendency observed in *M. lucifugus* was not tied to any bias in identifying the 5' regions using homology based methods or any bias in insertion preferences. However, specific conclusions cannot be drawn, since the mechanisms and conditions favoring the gene capture of *Helitrons* are not known yet. Another interesting feature observed is that captured gene fragments identified in bat *Helitrons* are in both sense (11 captures) and in antisense (13 captures) directions when compared to parental gene. Plant *Helitrons* were found to have a strong bias in capturing fragments in the sense direction even though most of the captures with antisense fragments were recent suggesting possible role of selection in eliminating antisense captures (Yang and Bennetzen 2009). The retention of *Helitrons* with gene captures in reverse orientation with respect to parental gene suggests a possible role in regulating the expression of those genes in the *M lucifugus* genome. Further experiments are needed to decipher how *Helitrons* have altered the gene expression in the little brown bat genome

3.4.3 Mechanism of *Helitron* gene captures

Irrespective of the high amplitude and frequency of the gene captures reported the mechanism of the gene capture by *Helitron* remains unknown. The proposed models for the gene captures are transposition starts at the 5' end and inefficient recognition of the 3' terminus

leads to the capture of flanking host sequence (Feschotte and Wessler 2001). The second model proposed that transposition starts at the 3' terminus and in efficient recognition of the 5' terminus leads to the capture of flanking host sequence. It is also proposed that gene capture is due to 'filler DNA' that was generated during double stranded breaks presumably during the *Helitrons* transposition (Kapitonov and Jurka 2007). Chimeric transposition was proposed when several truncated elements are present; transposition begins from one of the 3' terminus and identifies the 5' terminus of another *Helitron* mobilizing the host sequences in between (Tempel *et al.* 2007). The fifth model proposes that *Helitron* captures host genome sequences by recruiting a site specific recombinase similar to integrons (Lal *et al.* 2009). Upon closer inspection, we were able to identify the presence of full-length short *Helitron* (designated as 'progenitor element') at the 3' end of four *Helitrons* with gene captures, which provides support that transposition starts at the 3' end and missed the 5' end and captured upstream host sequences and terminated by another *Helitron* 5' end. The parental and copies of the captured fragments were examined for any evidence of a 'progenitor element' within 10kb (as maximum length of the *Helitron* observed in *M. lucifugus* genome is ~5kb). However, the progenitor *Helitron* was identified in the parental copy of TMBIM4 gene suggesting slippage of the termination signal might have lead to the capture of gene fragment (Figure 3.4). However, we were not able to identify *Helitrons* near the captured region (within ~10kb) in the other13 cases we analyzed suggesting either selection might have played a role in elimination of such progenitor elements or that they could be due to an alternative mechanisms of capturing host sequences.

3.4.4 Amplification of retrogenes: *Helitrons* have amplified five retrogenes

Generating gene duplicates have been recognized as an important mechanism in the evolution of genomic novelty (see review Zhang 2003). Several mechanisms are known to generate gene duplications including unequal crossing over, segment duplication (chromosomal or genome), transduction and retroposition (Zhang 2003). In our analysis, we identified

*Helitron*-mediated amplification (5-85copies) of five nested retrogenes. This reveals a novel mechanism for the generation of *Helitron*-mediated gene duplicates. No evidence of purifying selection were detected in those copies, however, further analysis is need to reveal their potential functions (Force *et al.* 1999, Zhang 2003).

Most of the retrogenes are pseudogenized because they do not carry regulatory sequences to drive their expression. How do functional retrogenes acquire promoters is an intriguing question. Retroposition of the RPLP0 gene has occurred in a *Helitron* containing the promoter of the SRPK1 gene. Even though the promoter is in the reverse orientation in the *Helitron*, it is provides an excellent example for how *Helitrons* could play a major role in promoter shuffling in addition to exon shuffling. Functional implications of these novel gene units and evolutionary implications are not known yet. However, these data also reveals the unique ability of *Helitrons* in carrying the retrogene without impacting its ability to transpose which could be further manipulated for making it a suitable vector for gene therapy.

3.4.5 *Helitrons*: drivers of vesper bat evolution

*M. lucifugus* belongs to Vespertilionidae family, (Order Chiroptera), the second most speciose mammal and geographically dispersed group of mammals (Nowack 1994). Other work from our lab has revealed that *Helitrons* are restricted to vesper bats and that *Helitrons* have invaded the genome horizontally, probably on two separate occasions (Thomas *et al.* 2010, Thomas *et al.* in review). Together with our discoveries presented herein we reveal that the HT of *Helitrons* has led to dramatic modifications of the *M. lucifugus* genome contributing to 11.5% of the DNA content and 37 novel gene combinations and further amplified (1-1145 copies) in the genome. Our understanding of the impact *Helitrons* have had on the evolutionary trajectory of vesper bats is in its infancy. It has been speculated that the amplifications of TEs in the ancestral vesper lineage might have triggered the rapid diversification of the lineage (Pritham and Feschotte 2007, Oliver and Greene 2009). The tremendous amplification of *Helitrons* along

with their propensity to shuffle exons and promoters, and even whole retrogenes could have potentially altered the genetic architecture of the bats and driven vesper bat evolution.

# APPENDIX A

## PERVASIVE HORIZONTAL TRANSFER OF ROLLING-CIRCLE TRANSPOSONS AMONG ANIMALS

Genome Biology and Evolution
SMBE
http://gbe.oxfordjournals.org



# Pervasive Horizontal Transfer of Rolling-Circle Transposons among Animals

Jainy Thomas, Sarah Schaack, and Ellen J. Pritham*

Department of Biology, University of Texas at Arlington

*Corresponding author: E-mail: pritham@uta.edu.

**Accepted:** 31 July 2010

## Abstract

Horizontal transfer (HT) of genes is known to be an important mechanism of genetic innovation, especially in prokaryotes. The impact of HT of transposable elements (TEs), however, has only recently begun to receive widespread attention and may be significant due to their mutagenic potential, inherent mobility, and abundance. *Helitrons*, also known as rolling-circle transposons, are a distinctive subclass of TE with a unique transposition mechanism. Here, we describe the first evidence for the repeated HT of four different families of *Helitrons* in an unprecedented array of organisms, including mammals, reptiles, fish, invertebrates, and insect viruses. The *Helitrons* present in these species have a patchy distribution and are closely related (80–98% sequence identity), despite the deep divergence times among hosts. Multiple lines of evidence indicate the extreme conservation of sequence identity is not due to selection, including the highly fragmented nature of the *Helitrons* identified and the lack of any signatures of selection at the nucleotide level. The presence of horizontally transferred *Helitrons* in insect viruses, in particular, suggests that this may represent a potential mechanism of transfer in some taxa. Unlike genes, *Helitrons* that have horizontally transferred into new host genomes can amplify, in some cases reaching up to several hundred copies and representing a substantial fraction of the genome. Because *Helitrons* are known to frequently capture and amplify gene fragments, HT of this unique group of DNA transposons could lead to horizontal gene transfer and incur dramatic shifts in the trajectory of genome evolution.

**Key words:** Helitrons, insect viruses, transposable elements, lateral transfer.

## Introduction

The movement of genetic material between reproductively isolated species, known as horizontal transfer (HT), is known to be an important process in genome evolution. In eukaryotes, this has been shown in the case of genes (for review, see Anderson 2005; Keeling and Palmer 2008) and, more recently, with transposable elements (TEs) (e.g., Silva et al. 2004; Casse et al. 2006; Diao et al. 2006; de Boer et al. 2007; Loreto et al. 2008; Pace et al. 2008; Bartolome et al. 2009; Roulin et al. 2009). TEs are mobile, parasitic pieces of genetic material that can mobilize and replicate within the host genome. Their inherent ability to replicate and integrate into the genome is likely to make them prone to HT (Kidwell 1992). HT has been proposed as an essential part of the lifecycle of some types of TEs in order to avoid co-evolved host suppression mechanisms aimed at limiting their mobility within lineages (Hartl et al. 1997; Silva et al. 2004). It has also been proposed that the propensity for HT could be related to the mechanism of transposition used (see Schaack, Gilbert and Feschotte 2010 for review). TEs are classified based on whether they move via an RNA intermediate (Class 1) or a DNA intermediate (Class 2), with further divisions based on the mechanism of integration (Wicker et al. 2007).

A unique group of rolling-circle (RC) DNA transposons called *Helitrons* (with atypical structural characteristics including 5′ TC and 3′ CTRR termini and a 16 to 20-nt palindrome upstream of the 3′ end [Feschotte and Wessler 2001; Kapitonov and Jurka 2001]) have been described in a wide array of eukaryotes including fungi (Poulter et al. 2003; Cultrone et al. 2007), plants (Kapitonov and Jurka 2001; Lal et al. 2003; Rensing et al. 2008; Yang and Bennetzen 2009a), insects (Kapitonov and Jurka 2001; Poulter et al. 2003; Langdon et al. 2009; Yang and Bennetzen 2009a; The International Aphid Genomics Consortium 2010), nematodes (Kapitonov and Jurka 2001), and vertebrates (Poulter et al. 2003; Zhou et al. 2006; Pritham and Feschotte 2007). In some cases, *Helitrons*

constitute a significant portion of the genomes (e.g., *Caenorhabiditis elegans*, *Arabidopsis thaliana*, *Myotis lucifugus*). *Helitrons*, unlike most other DNA transposons that use transposase, putatively encode a protein with a rolling circle initiator motif and *PIF1*-like DNA helicase domains and are categorized in their own subclass (Kapitonov and Jurka 2001; Wicker et al. 2007). Homology of the *Helitron*-encoded protein to bacterial RC transposons (IS91, IS1294, IS801), which are well known for their propensity to shuttle antibiotic resistance genes between distinct bacterial species (Toleman et al. 2006), reveals a distant relationship (Kapitonov and Jurka 2001). Like their bacterial cousins, some *Helitrons* function as "exon shuffling machines" (Feschotte and Wessler 2001). This ability is particularly pronounced in maize where it is estimated that at least 20,000 gene fragments have been picked up and shuffled by *Helitrons* (Du et al. 2009; Feschotte and Pritham 2009; Yang and Bennetzen 2009b). The ability to seize and recombine exons from multiple genes to create novel genetic units (Brunner et al. 2005; Gupta et al. 2005; Lal and Hannah 2005; Morgante et al. 2005; Xu and Messing 2006; Pritham and Feschotte 2007; Jameson et al. 2008; Langdon et al. 2009) makes HT of *Helitrons* especially intriguing because they can shuttle gene fragments between genomes.

This study expands our understanding of HT of TEs in several ways. First, we provide the first evidence for widespread, repeated HT of *Helitrons*, a distinctive group of transposons with a unique mechanism of replication. Second, in contrast to previous reports of widespread HT which have involved only *hAT* superfamily elements distributed largely among vertebrates (Pace et al. 2008; Gilbert et al. 2010), we show horizontally transferred *Helitrons* are frequently found in insect genomes. However, we have also identified cases of *Helitron* HT in vertebrates (bat, lizard, and jawless fish), a patchy distribution that indicates that certain host genomes are especially vulnerable to invasion. Third, this is the first report of *Helitron* HT in insect viruses, which could act as shuttle systems for the delivery of DNA between species (Loreto et al. 2008). Although HT has occasionally been invoked to explain discordant distributions in isolated cases (Kapitonov and Jurka 2003; Lal et al. 2009), our discovery of horizontally transferred *Helitrons* in viruses, insects, and vertebrates demonstrates the widest range of extensive HT among animals and possible vectors so far.

## Materials and Methods

*Helitrons* identified in *Myotis lucifugus* (the little brown bat) were used as an initial query (BlastN using default parameters (BlastN. . . [Altschul et al.1990]) to find *Helitrons* in other genomes available at the National Center for Biotechnology Information, including the whole genome shotgun, nucleotide collection (nr/nt), genome survey sequences, high throughput genomic sequences, and expressed sequence tag databases. Hits that were ≥65% identical to the query over >300 bp were examined and, when possible, full-length *Helitrons* were manually extracted. These elements were used as queries to find additional related *Helitrons*; the resulting hits were examined, and full-length *Helitrons* were extracted to generate a library of *Helitrons* for each species (details on all methods are in supplementary Materials and Methods, Supplementary Material online). *Helitrons* were then classified into families based on the following criteria according to Yang and Bennetzen (2009a, 2009b). We established conservative criteria to identify cases of HT that could be fully analyzed, including >80% identity at the 3′ end, a >400 bp portion of the internal region that is >80% identical, and divergence estimates among species that exclude the possibility of vertical inheritance (supplementary materials and methods, Supplementary Material online). *Helitrons* that share high levels of identity (>80%) from the same family in multiple species were aligned using MUSCLE (Edgar 2004) and analyzed as a group (including calculations of pairwise divergence [MEGA 4.0.2; Tamura et al. 2007], abundance [RepeatMasker version 3.2.7; A. F. A. Smit, R. Hubley, and P. Green, www.repeatmasker.org], and, when possible, calculations of amplification date estimates [as in Pritham and Feschotte 2007; Pace et al. 2008]).

## Results

### Identification, Classification, and Characterization of *Helitrons*

In a previous study, *Helitrons* were reported only in the little brown bat, *M. lucifugus*, among the 44+ publicly available mammalian genome sequences (Pritham and Feschotte 2007) that suggested the acquisition of these elements via HT. Because *M. lucifugus* is a good candidate for investigating possible HT, a deeper survey of *Helitrons* was performed, a previously uncharacterized family (*HeligloriaB_Ml*) was identified, and was used as a starting point for a series of Blast searches. These searches led to the subsequent identification of *Helitrons* from animals and animal viruses which were then classified into families based on their identity at the 3′ end (for family designation) and 5′ end (for subfamily designation), as in Yang and Bennetzen (2009a, 2009b; see Materials and Methods): the families were named *Heligloria*, *Helisimi*, *Heliminu*, and *Helianu*. Cases of recent HT were identified and analyzed when *Helitrons* of the same family that exhibited >80% identity at the 3′ end and contained a >400 bp portion of the internal region with >80% identity (see Materials and Methods) were found in diverged species (>35 million years ago [Ma]). *Helitrons* demonstrating high levels of identity that were inconsistent with vertical descent were found in many taxa, including insect viruses, many invertebrates (e.g., insects, nematodes, annelids, molluscs, and planaria), and vertebrates (e.g., salamanders, lizards, snakes, jawless fish, and bat; see supplementary table S1, Supplementary Material online). Those cases for

online). Fragments of copies of *Heliminu* and *Helianu* (>90% identical) are also present in a variety of other insects, including butterflies, moths, flies, and fleas (see supplementary table S1, Supplementary Material online). Paralogous or orthologous empty sites were identified for at least one member from each family to confirm the mobility of these elements (supplementary fig. S2, Supplementary Material online). The putative autonomous elements encode all the expected motifs and domains consistent with other described animal protein-coding *Helitrons* (Rep and helicase; supplementary fig. S3a, b, Supplementary Material online).

## Species-Specific Proliferation of *Helitrons* and Timing of Amplification

In the case of all four families, *Helitrons* have proliferated via amplification of nonautonomous copies. In the case of *HeligloriaB*, the autonomous partner responsible for the amplification of the non-autonomous elements was not identified in the genome sequences of bat, lizard, and insect. However, we were able to detect autonomous copies of *HeligoriaB* in the jawless fish genome sequence (supplementary table S2, Supplementary Material online). However, we were able to detect autonomous copies of *HeligloriaB* in jawless fish in the UCSC genome browser (supplementary table S2, Supplementary Material online). The discovery of autonomous partners for this family was likely hindered by low genome coverage and the older age of the family. It may be that with higher sequencing coverage or examination of additional genomes that the autonomous copies might be discovered.

Copy number varies across species but in some cases is high (up to 677 copies; table 1). Because we used the last 30 bp of the 3′ end, copy number estimates include all subfamilies. To estimate how much of the genome is occupied by each *Helitron* family, individual genomes were masked by the four families of *Helitrons* (not only the last 30 bp but with the entire element [table 1]). The apparent discrepancy in the copy number estimation and percent genome occupied is due to the difference in the methods employed. Some *Helitrons* tend to capture new 3′ ends, retaining the 5′ end and internal region. In those cases, copy number estimate (based on Blast with 3′ end) will be lower than the RepeatMasker estimate (based on the entire element). *Helitron* families appear to have differentially amplified or been retained in each host species (fig. 3), *Helisimi* is the most "successful," having amplified in *B. mori* to such an extent that it constitutes 0.2% of the genome and contributes almost 0.8 Mb of DNA (table 1). The timing of amplification of *HeligloriaB_Ml* in bat was estimated based on the average divergence of copies from the consensus sequence (3.8%) to be 14.1 Ma based on the neutral substitution rate as in Pace et al. (2008). In most of the cases, it was not possible to use this method because of difficulty reconstructing a consensus to estimate the ancestral copy and the lack of data on

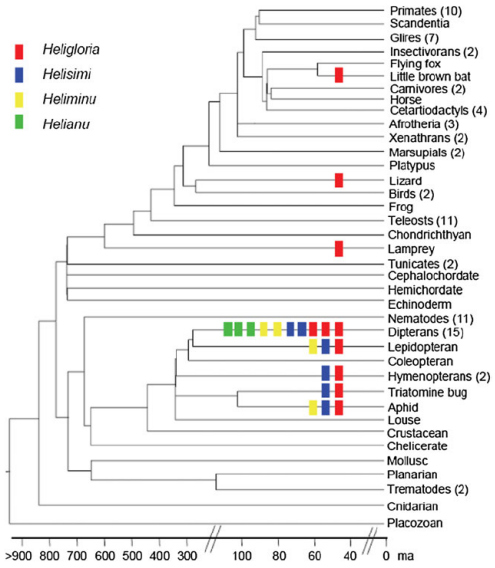



**Fig. 2.**—Schematic representation of phylogenetic relationships among animal lineages and estimated divergence times (Ma). Presence of horizontally transferred *Helitrons* from four different families in each lineage are denoted by rectangles (not placed relative to the timescale). Numbers in parentheses on the right indicate the number of species (when >1) for which whole genome sequence data are publicly available in the whole genome shotgun (National Center for Biotechnology Information).

mutation rates. In these cases, the percent divergence between a given *Helitron* copy (representative of a particular family) and its second-best hit (not with itself) were used as a proxy to estimate the relative timing of amplification (see supplementary table S4, Supplementary Material online). Even though *Helitrons* appear to be recently active in many genomes (≥99% identity between copies of some families in *R. prolixus*, *A. pisum*, and *B. mori*), there were other cases with no signs of recent activity (as low as 75% identity between copies).

## Evidence for HT

The high sequence identity (80–97%) of the *Helitrons* is not limited to the 5′ and 3′ ends but is also observed in the internal regions of all families (fig. 1*a* and *b* and supplementary table S3, Supplementary Material online). In many cases, the sequence identity of the *Helitrons* is exceptionally high compared with the divergence of the hosts (fig. 2). For example, there is 88% sequence identity between *Helitrons* in the mammal, *M. lucifugus*, and the lizard, which diverged 360 Ma and these diverged from the common ancestor of the jawless fish and the insect *R. prolixus* >600

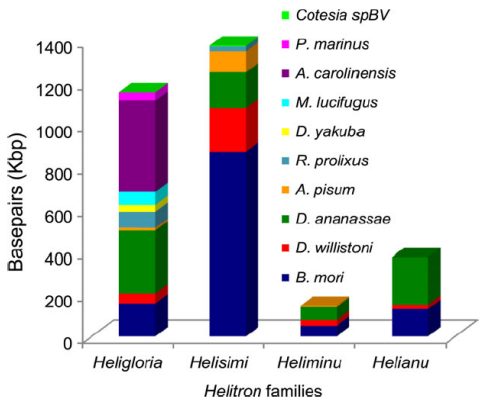



**Fig. 3.**—Distribution of *Helitron* families (*Heligloria*, *Helisimi*, *Heliminu*, and *Helianu*) across species and their contribution (shown in Kbp) toward the host genome.

and >750 Ma, respectively (fig. 2; Hedges et al. 2006). Similar patterns of sequence identity of *Helitrons* (86–97%) can be observed among insects of different orders (Lepidoptera, Diptera, Hemiptera) and the polydnaviruses inhabiting the hymenopteran parasitic wasps. The insects belonging to these orders diverged from their common ancestor >200 Ma (in the case of Diptera and Lepidoptera) and up to 350 Ma (in the case of Hemiptera). Previous work on TEs suggests that that these elements are not under host selective constraints (Silva and Kidwell 2000; Pace et al. 2008), and instead, TEs evolve neutrally upon inactivation of their transposition in the host genomes. The highly fragmented nature and lack of intact open reading frames of the *Helitrons* identified further supports the idea of lack of active transposition. The levels of divergence observed among *Helitrons* in these species are much lower than what would be expected based on direct estimates of neutral substitutions rates (e.g., $5.8 \times 10^{-8}$ mutations per site per year in *Drosophila* [Haag-Liautard et al. 2007]) given the current estimates of their divergence times (Hedges et al. 2006). Thus, HT is the best explanation for the exceedingly high sequence identity displayed by these TEs across widely diverged species. Another line of evidence that can be used to exclude the possibility of vertical transfer is the discontinuous presence of these elements across different species represented in the database. All four families of *Helitrons* have a patchy distribution with high sequence identity among vertebrates and insects (figs. 2 and 3). Although, it should be noted that false negative results might occur in genomes with low sequencing coverage and few copies. However, to attribute the patchy distribution observed here to vertical inheritance would require a nonparsimonious scenario of many cases of independent loss and intense activity in a small subset of lineages.

## Discussion

This is the first report of the HT of *Helitrons* among a diverse array of animal species. We identified 25 definitive cases of HT involving four families of *Helitrons* and nine animal species, including vertebrates and invertebrates that diverged, in some cases, more than 700 Ma (fig. 2 and table 1; for additional cases, see supplementary table S1, Supplementary Material online). Very high sequence identity among species (80–97%), in conjunction with the extremely fragmented nature of the *Helitrons* identified, preclude the possibility of vertical inheritance and selective constraint as an explanation for the similarity observed between elements across species. Our data reveal interesting patterns within the patchy distribution among animals, including the repeated invasion of some genomes by multiple *Helitron* families (figs. 2 and 3). Although some families (*Heliminu* and *Helianu*) are restricted to insects, *HeligloriaB* has invaded mammals, reptiles, and jawless fish, in addition to several insect species (table 1 and supplementary table S1, Supplementary Material online). Remarkably, two of the four *Helitron* families were also found in polydnaviruses that are involved in facilitating the parasitism of lepidopterans by hymenopteran wasps. We propose that the presence of *Helitrons* in viruses may reflect their role as vectors for HT between parasitic wasps and their hosts, although other routes of HT also likely exist.

### Mechanisms of Transfer

The remarkable breadth of species involved in these cases of HT (including not only bat, lizard, jawless fish but also triatomine bug, silkworm, aphid, drosophilids, and bracoviruses) suggests multiple mechanisms may underlie the horizontal spread of TEs. The identification of *Helitrons* in bracoviruses (double-stranded DNA viruses; Polydnaviridae family) is of particular interest as a potential vector for the delivery of TEs among species. These viruses have an obligatory relationship with parasitic wasps belonging to the Braconidae family, replicating only in wasp ovary cells and releasing fully formed viral particles during oviposition by the wasp into the lepidopteran larvae. The viral particles encode virulence factors that suppress the immunity of the lepidopteran (e.g., for review, see Webb et al. 2009), facilitating the growth of the wasp larvae. Yoshiyama et al. (2001) suggested that the close association between the parasitoid wasp and moth facilitates the HT of TEs, as in the case of the "mariner" element transferred between the braconid parasitoid wasp, *Ascogaster reticulatus*, and its moth host, the smaller tea tortrix, *Adoxophyes honmai*. There have been several reports of TE-like sequences in the genomes of DNA viruses (Miller DW and Miller LK 1982; Fraser et al. 1983; Fraser 1986; Friesen and Nissen 1990; Jehle et al. 1998; Drezen et al. 2006; Piskurek and Okada 2007; Desjardins et al. 2008; Marquez and Pritham

2010). If viruses shuttle TEs from one species to another, we might expect to see biased distributions of horizontally transferred TEs based on host susceptibility to a particular virus group. In fact, our data reveal biased distributions (e.g., *Helisimi* and *Heliminu* are only found in insects, whereas *HeligloriaB* is frequently found in vertebrates); however, the sampling bias of the available databases also influences our ability to detect patterns or identify mechanisms based on distribution.

In addition to viruses, some parasitic insects have also been implicated as agents of HT because of their intimate association with their hosts (e.g., Houck et al. 1991). Gilbert et al. (2010) recently found evidence for the HT of four DNA transposon families in *R. prolixus* and a wide array of tetrapods. Because *R. prolixus* is a sanguivorous parasite of mammals and vertebrates, transfer of DNA could occur through salivary deposition or blood intake by this species. The presence of closely related *Helitrons* in *R. prolixus* and *M. lucifugus*, a host of *R. prolixus*, further indicates this bug may be a candidate vector for transferring TEs. Other proposed mechanisms of transfer include endosymbiotic bacteria such as *Wolbachia* (Hotopp et al. 2007). It is known that *Wolbachia* infect *C. sesamiae* wasps (Mochiah et al. 2002), drosophilids, aphids (Jeyaprakash and Hoy 2000; The International Aphid Genomics Consortium 2010), *Rhodnius* sp. (Espino et al. 2009), and even nematodes (Fenn et al. 2006). In addition to the possibility of HT through *Wolbachia*, the bacteriophage of *Wolbachia* is also a potential vector for HT (Gavotte et al. 2007; Loreto et al. 2008). Additional experiments and taxon sampling are necessary to further delineate the role of host–parasite interactions and other intermediates such as bacteria and viruses, in the direction and frequency of HT of TEs and the as of yet unknown mechanisms underlying this process.

### Impact on Genome

Diverse mechanisms of HT can lead to recurrent invasions of genomes by *Helitrons*, thereby increasing the dynamic portion of the genome. The proposed rolling circle-like transposition mechanism could explain the tandem duplicates and arrays generated by *Helitrons* (supplementary fig. S4, Supplementary Material online, Pritham and Feschotte 2007; Schaack et al. 2010, Choi et al. 2010). The frequent capture of new 3′ and 5′ ends without disrupting their ability to transpose could extend the lifespan of *Helitrons* in the host genome and generate genetic diversity among elements. Their proposed replication mechanism also likely explains their unique propensity to capture host gene fragments, which could have a tremendous impact on the genome (e.g., Brunner et al. 2005; Gupta et al. 2005; Morgante et al. 2005; Xu and Messing 2006; Jameson et al. 2008; Du et al. 2009; Langdon et al. 2009; Yang and Bennetzen 2009b). Indeed, in *M. lucifugus*, *HelibatN3* has captured the promoter and first exon of the NUBPL (a single copy gene which is highly conserved in mammals) and amplified it to high copy number (>1,000; Pritham and Feschotte 2007). Amplification is thought to closely follow invasion of a naive genome (Pace et al. 2008) and results in opportunities for genetic innovation. Genetic innovation, in turn, leads to diversification within the lineage, a possibility supported by the occurrence of multiple waves of TE invasion in the bat lineage around the time of their rapid diversification, 16–25 Ma (Teeling et al. 2005; Pritham and Feschotte 2007; Ray et al. 2008; Oliver and Greene 2009; Zeh et al. 2009; Gilbert et al. 2010). We conclude that the HT, colonization, and amplification of *Helitrons* are rampant and widespread across animals and can play a major role in genome evolution.

## Supplementary Material

Supplementary dataset, materials and methods, figures S1–S5, and tables are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

This work was supported by start-up funds from the University of Texas-Arlington to E.J.P. and National Science Foundation award 0805546 to S.S. We would like to acknowledge the genome sequencing consortiums for sequencing the *M. lucifugus*, *A. carolinensis*, *P. marinus*, and *R. prolixus* and *C.sesamiae* bracovirus genomes. We would also like to thank Brian Fontenot, Matt Carrigan, and two anonymous reviewers for helpful comments on the manuscript.

## Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Anderson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci. 62:1182–1197.

Bartolome C, Bello X, Maside X. 2009. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. Genome Biol. 10:2.

Brunner 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.

Casse N, Bui QT, Nicolas V, Renault S, Bigot Y, Laulier M. 2006. Species sympatry and horizontal transfers of Mariner transposons in marine crustacean genomes. Mol Phylogenet Evol. 40:609–619.

Cultrone A, Dominguez YR, Drevet C, Scazzocchio C, Fernandez-Martin R. 2007. The tightly regulated promoter of the *xanA* gene of *Aspergillus nidulans* is included in a *Helitron*. Mol Microbiol. 63:1577–1587.

de Boer JG, Yazawa R, Davidson WS, Koop BF. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics. 8:442.

Desjardins CA, et al. 2008. Comparative genomics of mutualistic viruses of *Glyptapanteles* parasitic wasps. Genome Biol. 9:12.

Diao X, Freeling M, Lisch D. 2006. Horizontal transfer of a plant transposon. PLoS Biol. 4:e5.

Drezen JM, et al. 2006. The few virus-like genes of *Cotesia congregata* bracovirus. Arch Insect Biochem Physiol. 61:110–122.

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.

Dupuy C, Huguet E, Drezen JM. 2006. Unfolding the evolutionary story of polydnaviruses. Virus Res. 117:81–89.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Espino CI, et al. 2009. Detection of *Wolbachia* bacteria in multiple organs and feces of the triatomine insect *Rhodnius pallescens* Hemiptera, Reduviidae. App Environ Microbiol. 75:547–550.

Fenn K, et al. 2006. Phylogenetic relationships of the Wolbachia of Nematodes and Arthropods. PLoS Pathog. 2(10):e94.

Feschotte C, Pritham EJ. 2009. A cornucopia of *Helitrons* shapes the maize genome. Proc Natl Acad Sci U S A. 106:19747–19748.

Feschotte C, Wessler SR. 2001. Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. Proc Natl Acad Sci U S A. 98:8923–8924.

Fraser MJ 1986. Transposon-mediated mutagenesis of baculoviruses: transposon shuttling and implications for speciation. Ann Entomol Soc Am. 79:773–783.

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. 47:287–300.

Friesen PD, Nissen MS. 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. Mol Cell Biol. 10:3067–3077.

Gavotte L, et al. 2007. A survey of the bacteriophage WO in the endosymbiotic bacteria Wolbachia. Mol Biol Evol. 24:427–435.

Gilbert C, Schaack S, Pace JK II, Brindley PJ, Feschotte C. 2010. A role for host–parasite interactions in the horizontal transfer of transposons across phyla. Nature. 464:1347–1350.

Gupta S, Gallavotti A, Stryker GA, Schmidt RJ, Lal SK. 2005. A novel class of *Helitron*-related transposable elements in maize contain portions of multiple pseudogenes. Plant Mol Biol. 57:115–127.

Haag-Liautard C, et al. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature. 445:82–85.

Hartl DL, et al. 1997. What restricts the activity of mariner-like transposable elements? Trends Genet. 13:197–201.

Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 22: 2971–2972.

Hotopp JCD, et al. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science. 317:1753–1756.

Houck MA, et al. 1991. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science. 253:1125–1129.

The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid acyrthosiphon pisum. PLoS Biol. 82:e1000313.

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

Jehle JA, Nickel A, Vlak JM, Backhaus H. 1998. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. J Mol Evol. 46:215–224.

Jeyaprakash A, Hoy MA. 2000. Long PCR improves *Wolbachia* DNA amplification: WSP sequence found in 76% of sixty-three arthropod species. Insect Mol Biol. 9:393–405.

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

Kapitonov V V, Jurka J. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc Natl Acad Sci U S A. 100:6569–6574.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9:605–618.

Kidwell MG. 1992. Horizontal transfer of *P*-elements and other short inverted repeat transposons. Genetica. 86:275–286.

Lal SK, Giroux MJ, Brendel V, Vallejos CE, Hannah LC. 2003. The maize genome contains a *Helitron* insertion. Plant Cell. 15:381–391.

Lal SK, Hannah LC. 2005. *Helitrons* contribute to the lack of gene colinearity observed in modern maize inbreds. Proc Natl Acad Sci U S A. 102:9993–9994.

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

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

Loreto ELS, Carareto CMA, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity. 100:545–554.

Marquez CP, Pritham EJ. 2010. Phantom, a new subclass of Mutator DNA transposons found in insect viruses and widely distributed in animals. Genetics. 185:1507–1517.

Miller DW, Miller LK. 1982. A virus mutant with an insertion of a *Copia*-like transposable element. Nature. 299:562–564.

Mochiah MB, Ngi-Song AJ, Overholt WA, Stouthamer R. 2002. *Wolbachia* infection in *Cotesia sesamiae* (Hymenoptera: Braconidae) causes cytoplasmic incompatibility: implications for biological control. Biological Control. 25:74–80.

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

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

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.

Piskurek O, Okada N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. Proc Natl Acad Sci U S A. 104(29):12046–12051.

Poulter RTM, Goodwin TJD, Butler MI. 2003. Vertebrate helentrons and other novel *Helitrons*. Gene. 313:201–212.

Pritham 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.

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

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

Roulin A, et al. 2009. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae. BMC Evol Biol. 9:58.

Schaack S, Choi E, Lynch M, Pritham EJ. 2010. DNA transposons and the role of recombination in mutation accumulation in *Daphnia pulex*. Genome Biol. 11:R46.

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

Silva JC, Kidwell MG. 2000. Horizontal transfer and selection in the evolution of *P* elements. Mol Biol Evol. 17:1542–1557.

Silva JC, Loreto EL, Clark JB. 2004. Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol. 6:57–71.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.
Teeling EC, et al. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. Science. 307:580–584.
Toleman MA, Bennett PM, Walsh TR. 2006. ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev. 70:296–316.
Webb B, Fisher T, Nusawardani T. 2009. The natural genetic engineering of polydnaviruses. Ann N Y Acad Sci. 1178:146–156.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.
Xu JH, Messing J. 2006. Maize haplotype with a helitron-amplified cytidine deaminase gene copy. BMC Genet. 7:52.
Yang LX, Bennetzen JL. 2009a. Structure-based discovery and description of plant and animal *Helitrons*. Proc Natl Acad Sci U S A. 106:12832–12837.
Yang LX, Bennetzen JL. 2009b. Distribution, diversity, evolution, and survival of *Helitrons* in the maize genome. Proc Natl Acad Sci U S A. 106:19922–19927.
Yoshiyama M, et al. 2001. Possible horizontal transfer of a transposable element from host to parasitoid. Mol Biol Evol. 18:1952–1958.
Zeh DW, Zeh JA, Ishida Y. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. Bioessays. 31:715–726.
Zhou Q, et al. 2006. Helitron transposons on the sex chromosomes of the platyfish *Xiphophorous maculates* and their evolution in animal genomes. Zebrafish. 31:39–52.

**Associate editor:** Ross Hardison

APPENDIX B

IDENTIFICATION NUMBER, FAMILY, NAME OF DIFFERENT BAT TISSUE OR DNA SAMPLES OBTAINED FROM MUSEUMS AND FROM PERSONAL COLLECTION.

List of bat samples obtained from Museum of Texas Tech University, Texas.

| TK NUMBER | FAMILY | NAME OF THE BAT SAMPLES |
|---|---|---|
| 152133 | Hipposideridae | *Hipposideros larvatus* |
| 18701 | Noctilionidae | *Noctillio leporinus* |
| 134826 | Molossidae | *Eumops wilsoni* |
| 18818 | Thyropteridae | *Thyroptera tricolor* |
| 152137 | Megadermatidae | *Megaderma spasma* |
| 152117 | Pteropodidae | *Balionycteris maculata* |
| 152047 | Pteropodidae | *Macroglossus* sp |
| 152238 | Emballonuridae | *Emballonura alecto* |
| 152256 | Rhinolophidae | *Rhinolophus sp* |
| 101001 | Phyllostomidae | *Sturnira sp* |
| 134970 | Phyllostomidae | *Vampyrum spectrum* |
| 117667 | Phyllostomidae | *Brachyphylla cavernarum* |
| 18826 | Phyllostomidae | *Desmodus rotundus* |
| 27682 | Phyllostomidae | *Artibeus jamaicensis* |
| 136092 | Phyllostomidae | *Carollia sp.* |
| 34863 | Phyllostomidae | *Anoura geoffroy* |
| 101015 | Phyllostomidae | *Sturnira ludovici* |
| 134789 | Phyllostomidae | *Sturnira lilium* |
| 117665 | Phyllostomidae | *Monophyllus plethodon* |
| 104135 | Phyllostomidae | *Rhinophylla sp* |
| 135710 | Phyllostomidae | *Rhinophylla alethina* |
| 134597 | Phyllostomidae | *Lophostoma silvicola* |
| 104517 | Phyllostomidae | *Micronycteris megalotis* |

| | | |
|---|---|---|
| 104603 | Phyllostomidae | *Uroderma bilobatum* |
| 135117 | Phyllostomidae | *Phyllostomus discolor* |
| 104624 | Phyllostomidae | *Phyllostomus elongatus* |
| 135927 | Phyllostomidae | *Lonchophylla concava* |
| 135803 | Phyllostomidae | *Lonchophylla thomasi* |
| 104313 | Phyllostomidae | *Carollia sp* |
| 152268 | Vespertilionidae | *Hesperoptenus tomesi* |
| 134649 | Vespertilionidae | *Lasiurus sp* |
| 104003 | Vespertilionidae | *Myotis sp* |
| 18815 | Vespertilionidae | *Eptesicus furinalis* |
| 104641 | Vespertilionidae | *Myotis sp* |
| 135013 | Vespertilionidae | *Myotis sp* |
| 152040 | Vespertilionidae | *Myotis ridelyi* |
| 152265 | Vespertilionidae | *Arielulus cuprosus* |
| 152059 | Vespertilionidae | *Murina rosendali* |
| 152074 | Vespertilionidae | *Glischropus tylopus* |
| 152082 | Vespertilionidae | *Tylonycteris robustula* |
| 152087 | Miniopteridae | *Miniopterus maganater* |

Note. Cytochrome b analysis of the bat samples used in the study was done to confirm the identity and the best hit (blastn analysis to the nr/nt database) is given in Appendix C

List of bat samples provided by American Museum of Natural History and Department of Mammalogy, New York

| DEPT/PARTNER ID | AMCC ID# | FAMILY | BAT SAMPLE |
|---|---|---|---|
| AMNH 273176 | 109612 | Phyllostomidae | *Trachops cirrhosus* |
| AMNH 274571 | 109312 | Mormoopidae | *Pteronotus davyi* |
| AMNH 269115 | 110395 | Mormoopidae | *Pteronotus sp* |
| AMNH 274632 | 102719 | Mormoopidae | *Pteronotus macleayii* |
| AMNH 275500 | 103036 | Mormoopidae | *Pteronotus quadridens* |
| AMNH265974 | 110378 | Noctilionidae | *Noctilio leporinus* |
| AMNH273085 | 109543 | Noctilionidae | *Noctilio albiventris* |
| AMNH272707 | 109673 | Vespertilionidae | *Myotis* sp |
| AMNH275501 | 103037 | Mormoopidae | *Mormoops blainvilli* |
| AMNH273155 | 110107 | Thyropteridae | *Thyroptera tricolor* |
| AMNH272742 | 109704 | Vespertilionidae | *Myotis sp* |
| AMNH268594 | 110458 | Molossidae | *Eumops* sp |
| AMNH268561 | 110295 | Phyllostomidae | *Sturnira tildae* |
| MUSM15277 | 109578 | Phyllostomidae | *Anoura* sp |
| AMNH 274576 | 109317 | Phyllostomidae | *Artibeus jamaicens* |
| AMNH272151 | 110783 | Rhinolophidae | *Rhinolophus* sp |
| AMNH267504 | 110497 | Phyllostomidae | *Desmodus rotundus* |
| AMNH272137 | 110778 | Megadermatidae | *Megaderma lyra* |
| MUSM13231 | 109708 | Phyllostomidae | *Phylloderma stenops* |
| AMNH 730575 | 109531 | Phyllostomidae | *Macrophyllum macrophyllum* |
| AMNH272746 | 109709 | Phyllostomidae | *Lophostoma silvicola* |

Bat samples provided by Dr. David Ray, University of Mississippi

| FAMILY | BAT SAMPLES |
|---|---|
| Vespertilionidae | *Nycticieus humeralis* |
| Vespertilionidae | *Lasiurus* sp |
| Vespertilionidae | *Pipistrellus subflavus* |
| Vespertilionidae | *Myotis sp* |

APPENDIX C

CYTOCHROME B ANALYSIS OF THE DIFFERENT BAT DNA/ TISSUE SAMPLES USED FOR THE STUDY

| Sample ID | Name of the bat | Best hit (nr/nt database at NCBI) | # bp | % ID |
|---|---|---|---|---|
| TK152087 | *Miniopterus magnater* | *Miniopterus magnater* | 1140 | 99 |
| TK136092 | *Carollia sp.* | *Carollia sowelli* | 1120 | 99 |
| 109317 | *Artibeus jamaicensis* | *Artibeus jamaicensis* | 1140 | 99 |
| 110497 | *Desmodus rotundus* | *Desmodus rotundus* | 1128 | 94 |
| 103037 | *Mormoops blainvilli* | *Mormoops blainvilli* | 1095 | 99 |
| 109312 | *Pteronotus davyi* | *Pteronotus davyi* | 1140 | 99 |
| 110395 | *Pteronotus sp* | *Pteronotus rubiginosus* | 1140 | 99 |
| 103036 | *Pteronotus quadridens* | *Pteronotus quadridens* | 1140 | 100 |
| 110295 | *Sturnira tildae* | *Sturnira tildae* | 767 | 99 |
| 109709 | *Lophostoma silvicola* | *Lophostoma silvicola* | 1140 | 96 |
| 109531 | *Macrophyllum macrophyllum* | *Macrophyllum macrophyllum* | 1140 | 97 |
| 109673 | *Myotis sp* | *Myotis albescens* | 1056 | 99 |
| TK135803 | *Lonchophylla thomasi* | *Lonchophylla thomasi* | 1018 | 90 |
| 109704 | *Myotis sp* | *Myotis simus* | 1140 | 98 |
| 102719 | *Pteronotus macleayii* | *Pteronotus macleayii* | 1140 | 100 |
| 134789 | *Sturnira lilium* | *Sturnira lilium* | 762 | 98 |
| 109612 | *Trachops cirrhosus* | *Trachops cirrhosus* | 1125 | 93 |
| 110107 | *Thyroptera tricolor* | *Thyroptera tricolor* | 1140 | 100 |
| 109578 | *Anoura sp* | *Anoura geoffroyi* | 1132 | 86 |
| 110458 | *Eumops auripendulus** | *Eumops perotis* | 392 | 82 |
| 104517 | *Micronycteris megalotis* | *Micronycteris megalotis* | 350 | 98 |
| TK117665 | *Monophyllus plethodon* | *Monophyllus plethodon* | 798 | 99 |
| TK134970 | *Vampyrum spectrum* | *Vampyrum spectrum* | 370 | 89 |
| TK135927 | *Lonchophylla concava** | *Lonchophylla mordax* | 746 | 98 |
| TK117667 | *Brachyphylla cavernarum* | *Brachyphylla cavernarum* | 737 | 99 |
| TK34863 | *Anoura geoffroyi* | *Anoura geoffroyi* | 863 | 99 |
| TK152265 | *Arielulus cuprosus** | *Eptesicus diminutus* | 767 | 83 |
| TK152117 | *Balionycteris maculata* | *Balionycteris maculata* | 728 | 98 |
| TK104313 | *Carollia sp* | *Carollia perspicillata* | 893 | 99 |
| TK18815 | *Eptesicus furnalis* | *Eptesicus furinalis* | 731 | 98 |
| TK152074 | *Gischropus tylopus** | *Pipistrellus abramus* | 952 | 83 |
| TK152268 | *Hesperoptenus tomesi** | *Rhogeessa velilla* | 739 | 83 |
| TK152133 | *Hipposideros larvatus* | *Hipposideros larvatus* | 927 | 95 |
| TK134649 | *Lasiurus sp.* | *Lasiurus xanthinus* | 797 | 85 |
| 109709 | *Lophostoma silvicola* | *Lophostoma silvicola* | 1140 | 96 |
|  | *Lasiurus seminolis** | *Lasiurus sp* | 778 | 82 |
| 109543 | *Noctilio albiventris* | *Noctilio albiventris* | 880 | 99 |
| TK152040 | *Myotis ridleyi** | *Myotis ikonnikovi* | 973 | 87 |

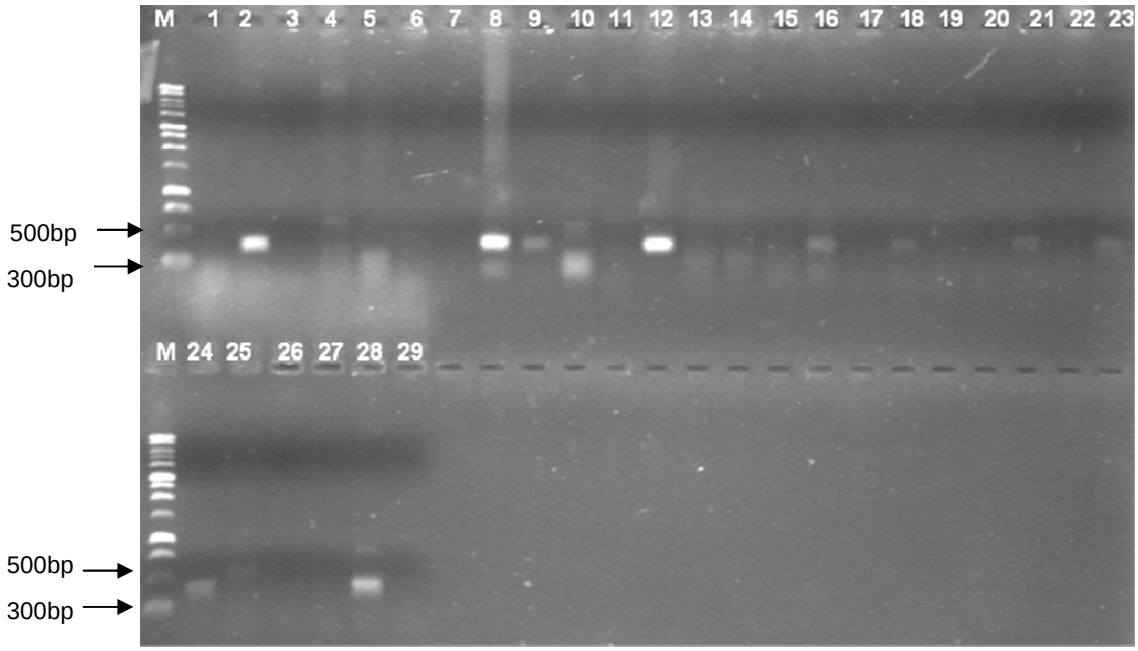
| | | | | |
|---|---|---|---|---|
| TK152059 | *Murina rosendali** | *Murina suilla* | 863 | 99 |
| TK152047 | *Macroglossus sp* | *Macroglossus minimus* | 718 | 89 |
| | *Pipistrellus subflavus* | *Pipistrellus subflavus* | 842 | 99 |
| | *Myotis sp* | *Myotis auriculus* | 854 | 98 |
| 110778 | *Megaderma lyra* | *Megaderma lyra* | 802 | 97 |
| | *Nycticeius humeralis* | *Nycticeius humeralis* | 501 | 99 |
| TK135710 | *Rhinophylla alethina* | *Rhinophylla alethina* | 897 | 99 |
| TK152256 | *Rhinolophus sp* | *Rhinolophus macrotis* | 921 | 90 |
| TK104135 | *Rhinophylla sp* | *Rhinophylla fischerae* | 834 | 96 |
| TK101015 | *Sturnira ludovici* | *Sturnira ludovici* | 751 | 99 |
| TK152082 | *Tylonycteris robustula** | *Tylonycteris pachypus* | 954 | 87 |
| TK104603 | *Uroderma bilobatum* | *Uroderma bilobatum* | 662 | 96 |
| TK104624 | *Phyllostomus elongatus** | *Phyllostomus hastatus* | 660 | 90 |
| TK18701 | *Noctillio leporinus* | *Noctillio leporinus* | 1066 | 99 |
| TK152238 | *Emballonura alecto* | *Emballonura alecto* | 402 | 99 |
| TK101001 | *Sturnira sp* | *Sturnira lilum* | 1140 | 94 |
| TK134826 | *Eumops wilsoni** | *Eumops sp.* | 646 | 87 |
| TK104003 | *Myotis sp* | *Myotis atacamensis* | 967 | 91 |
| TK135013 | *Myotis sp* | *Myotis atacamensis* | 816 | 91 |
| TK104641 | *Myotis sp* | *Myotis atacamensis* | 843 | 91 |
| 109708 | *Phylloderma stenops* | *Phylloderma stenops* | 689 | 90 |
| TK135117 | *Phyllostomus discolor** | *Phyllostomus hastatus* | 757 | 78 |
| TK152137 | *Megaderma spasma* | *Megaderma spasma* | 369 | 83 |
| 110783 | *Rhinolophus sp* | *Rhinolophus sp* | 300 | 94 |
| TK104158 | *Lophostoma sp* | *Lophostoma evotis* | 902 | 97 |

Note:

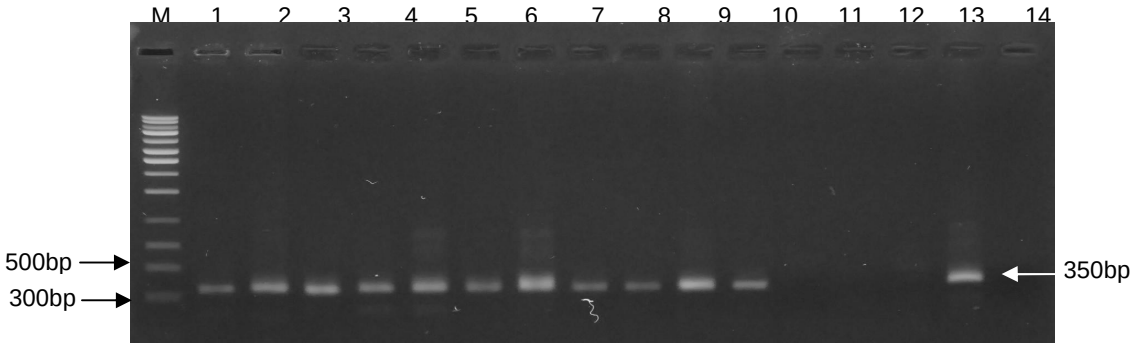
To confirm the identity of the source species of the tissue and DNA samples we systematically sequenced the cytob gene. If the best hit to cytob was congruent to the species name provided, we considered it a match and the name in column c will match column d. If the best hit was to another species and the species named was in the database, we considered it a failed match. If the % sequence identity was >90% and was to a species in the same genus, we entered the genus name sp in column c. If no cytob data has been deposited in the database at NCBI for any species in the genus, and our cytob data revealed significant hits (80-89% identity) to other members of the same family than we call them by the museum ID. An asterisk indicates absence of cytob data for that species in the database^. Cytob gene was amplified using the primers from Hoffman and Baker 2001.

# APPENDIX D

## AMPLIFICATION OF *HELITRONS* IN DIFFERENT ANIMALS AND BATS USING DEGENERATE PRIMERS



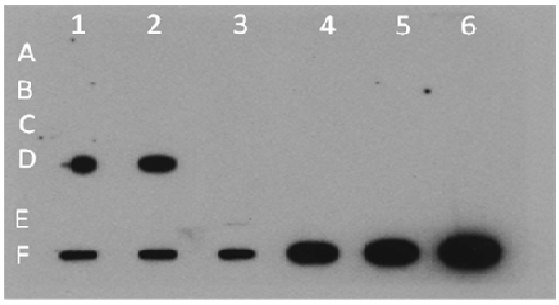

Amplification of *Helitrons* (350 bp) in different animals using the degenerate primers. Electrophoresis was done using 1% agarose gel at 100 V and the DNA samples used were. 1.*Typhlomolge rathbuni* (Salamander),2.*Thamnophis proximus* (Colubrid snake)3. *Dromaius novaehollandiae* (Emu)4. *Eurycea* (Salamander) 5. *Cactua sanguine* (parrot) 6. *Morelia* (python), 7.*Plethondar serratus* (salamander) 8.*Batrachoseps attenuates* (salamander) 9. *Abrouia fimbriata* (lizard),10. Dipsadine snake 11. *Dermophis mexianus* (Amphibian) 12. *Bombina orientalis* (Toad) 13. *Ambystoma opacum* (salamander)14. *Amphiuma pholeter* (salamander)15. *Rhyacotriton kezeri* (salamander) 16. *Psudobranchus axanthus* (salamander) 17. *Desonognathus guadramaculatus* (dusky salamander) 18. *Gonatodes albogularis* (Gecko)19. *Gonatodes falcoreusis* (Gecko) 20.*Gonatodes humeralis* (Gecko) 21.*Gonatodes manessi* (Gecko) 22. *Lepidopharis xanthostiguna* (Gecko) 23. *Sphaerodactytus molei* (Gecko), 24. *Coleodactylus amazonicus* (Gecko)25. Killerfish 26. Horseshoe crab 27. *Ciona* 28. *Anolis carolinensis* 29. Negative



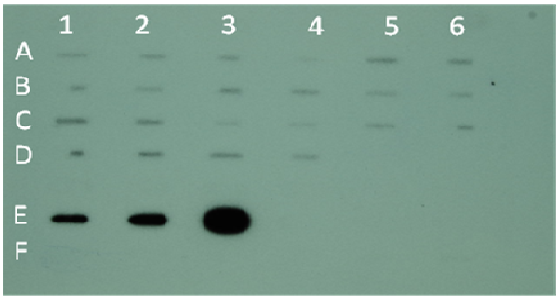

Amplification of *Helitrons* (350 bp) in different bats using the degenerate primers. The different bat species used are 1.*Myotis sp.* 2. *Lasiurus* sp 3. *Hesperoptenus tomesi*, 4. *Myotis* sp. 5. *Nycticeius humeralis* 6. *Murina rosendali* 7. *Pipistrellus subflavus* 8. *Lasirus* sp. 9. *Arielulus cuprosus*, 10. *Glischropus tylopus*, 11. *Tylonycteris robustula* 12. *Emballonura alecto* (Emballonuridae) 13. *Miniopterus magnater* (Miniopteridae) 14. *Balionycteris maculata* (Pteropodidae)15. *Myotis* sp 16. Negative M. 1kb ladder (Bat samples loaded in Lane 1-11 and 15 belongs vespertilionidae; Electrophoresis was done using 1% agarose gel at 100 V)

# APPENDIX E

## GENOMIC SLOT BLOT AND SOUTHERN BLOT ANALYSIS TO DETECT THE PRESENCE OF *HELITRONS* IN BATS




a) Autoradiograph after hybridizing with *Helibat* probe

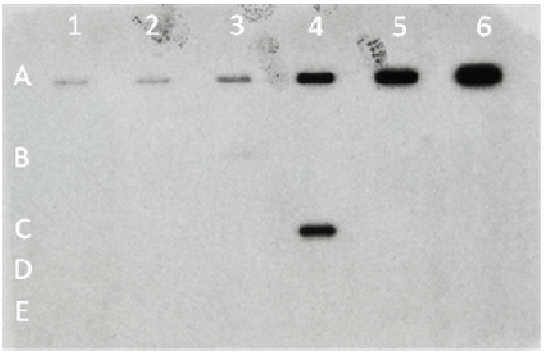



b) Autoradiograph after hybridizing with Rag1 probe

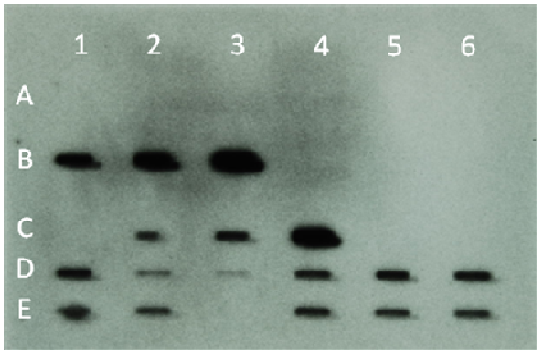
c) Map of the DNA samples loaded into the membrane

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | *Rhinolophus* | *M. lyra* | *N. albiventris* | *N. leporinus* | *P. davyi* | *P. macleayi* |
| B | *Pteronotus* sp | *P. quadridens* | *M. blainvilli* | *D. rotundus* | *T. cirrhosus* | *M. macrophyllum* |
| C | *L. silivicola* | *P. stenops* | *A. geoffroy* | *T. tricolor* | *S. tilidae* | *A. jamaicens* |
| D | *Myotis* sp. | *Myotis* sp. | *E.auripendulus* | *Mus musculus* | | |
| E | 0.4(199) | 1(496) | 2.5(1241) | | | |
| F | 0.4(199) | 1(496) | 2.5(1241) | 10(4964) | 20(7447) | 30(14894) |

Genomic Slot Blot analysis. a) autoradiograph showing the hybridization of the nylon membrane with the *Helibat* probe. b) autoradiograph showing the hybridization of the membrane with the Rag1 probe. c) Map of the DNA samples (1µg) loaded into the membrane and the combination of letters and numbers in the radiograph corresponds to the DNA samples loaded in the membrane. E and F lanes are loaded with the dilutions (ng) of the plasmid DNA containing the *Helitron* and Rag1 PCR products respectively and the numbers within the parenthesis indicates the copy number estimated by comparing the number of bat genomes (estimated genome size 2.3GB) and plasmid genomes (4.3kb) loaded onto the membrane. A, B, C, and D corresponds to the different bat species DNA and the mouse DNA (negative control) loaded. Details regarding the bat samples can be found in Supplemental Table 2.1




a) Autoradiograph after hybridizing with *Helibat* probe

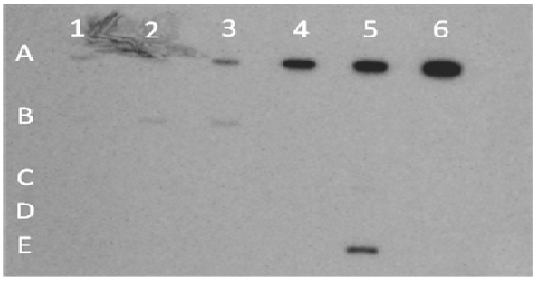



b) Autoradiograph after hybridizing with Rag1 probe

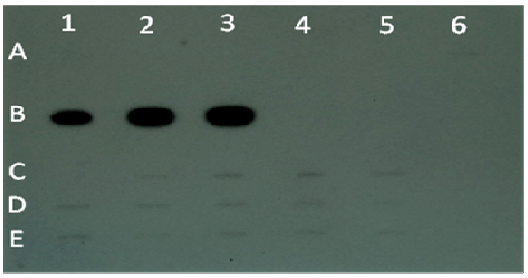
c) Map of the DNA samples loaded into the membrane

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | 0.3(149) | 1(496) | 2.5(1241) | 10(4964) | 20(7447) | 25(12412) |
| B | 0.5 (248) | 1(496) | 2.5(1241) | | | |
| C | | *T. cirrhosus* | *M. plethodon* | *Myotis sp* | | |
| D | *M. megalottis* | *Carollia* | *Mus musculus* | *Phyllostomus* sp | *L. thomasi* | *L. silivicola* |
| E | *V.spectrum* | *Lonchophylla* | *Macroglossus* | *B. cavernarum* | *S. lilium* | *R. alethina* |

**Genomic Slot Blot analysis**. a) autoradiograph showing the hybridization of the nylon membrane with the radio-labeled *Helibat* probe. b) autoradiograph showing the hybridization of the nylon membrane with the radio-labeled Rag1 probe. c) Map of the DNA samples (1µg) loaded into the membrane and the combination of letters and numbers in the radiograph corresponds to the DNA samples loaded in the membrane. A and B lanes are loaded with the dilutions (ng) of the plasmid DNA containing the *Helitron* and Rag1 PCR products respectively and the numbers within the parenthesis indicates the copy number estimated by comparing the number of bat genomes (estimated genome size 2.3GB) and plasmid genomes (4.3kb) loaded onto the membrane. C, D and E correspond to the different bat species DNA and the mouse DNA (negative control) loaded. Details regarding the bat samples can be found in Supplemental Table 2.1.




a) Autoradiograph after hybridizing with *Helibat* probe

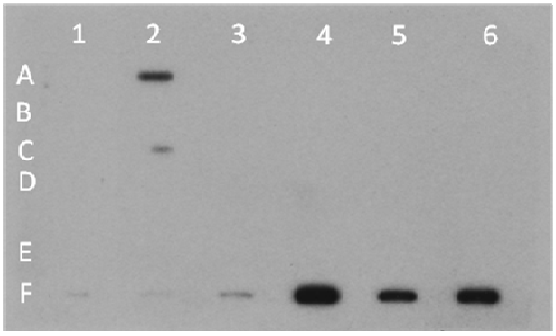



b) Autoradiograph after hybridizing with Rag1 probe

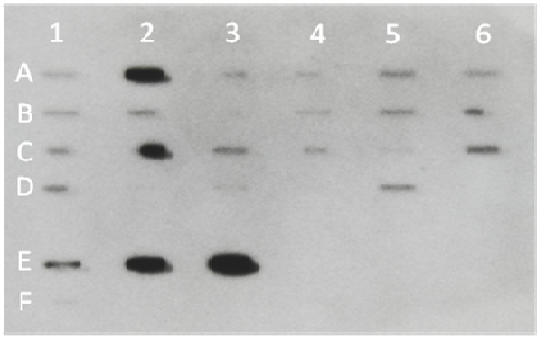
c) Map of the DNA samples loaded into the membrane

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | 0.5 (248) | 1(496) | 2.5(1241) | 10(4964) | 20(7447) | 25(12412) |
| B | 0.5 (248) | 1(496) | 2.5(1241) | | | |
| C | | *Carollia* | *M. megalottis* | *T.cirrhosus* | *Mus musculus* | |
| D | *Vampyrum* | *L silivicolum* | *P.discolor* | *B.cavernarum* | *M.plethodon* | |
| E | *L. thomasi* | *Lonchophylla* sp | *R. alethina* | *Sturnira lilium* | ***Lasiurus* sp** | |

Genomic Slot Blot analysis. a) autoradiograph showing the hybridization of the membrane with the helibat probe. b) autoradiograph showing the hybridization of the membrane with the Rag1 probe. c) Map of the DNA samples (1µg) loaded into the membrane and the combination of letters and numbers in the radiograph corresponds to the DNA samples loaded in the membrane. A and B lanes are loaded with the dilutions (ng) of the plasmid DNA containing the *Helitron* and Rag1 PCR products respectively and the numbers within the parenthesis indicates the copy number estimated by comparing the number of bat genomes (estimated genome size 2.3GB) and plasmid genomes (4.3kb) loaded onto the membrane. C, D and E correspond to the different bat species DNA and the mouse DNA (negative control) loaded. Details regarding the bat samples can be found in Supplemental Table 2.1.




a) Autoradiograph after hybridizing with *Helibat* probe

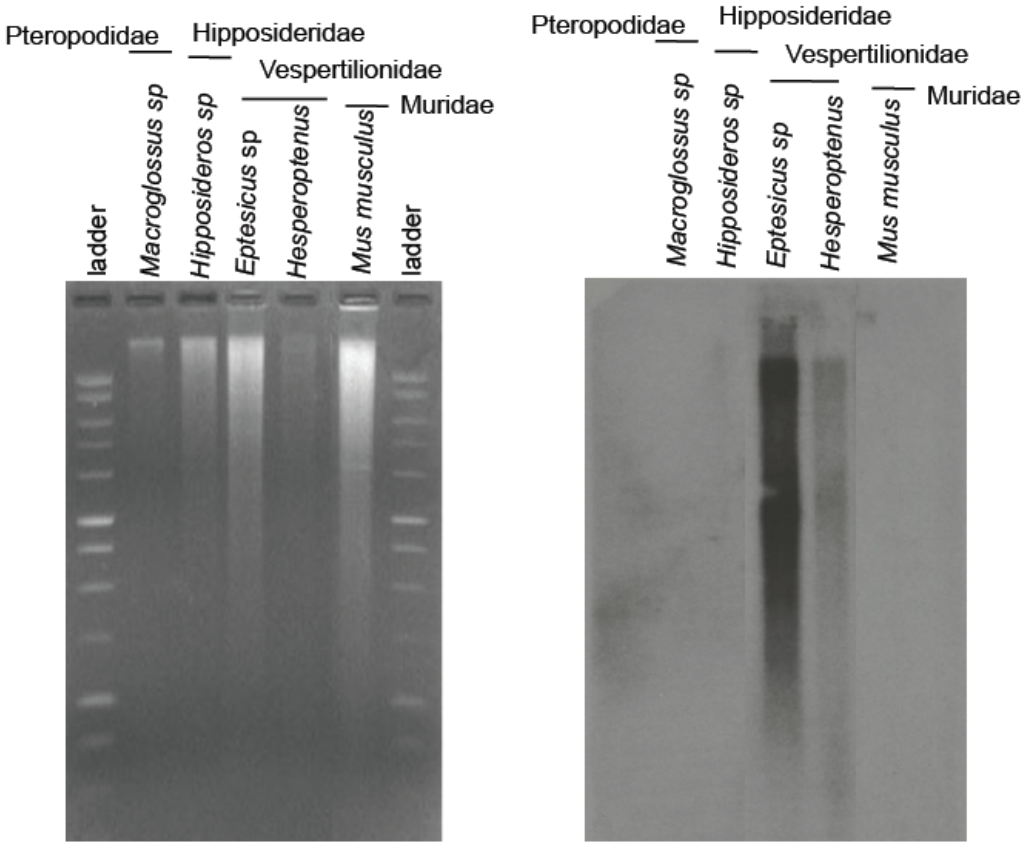



b) Autoradiograph after hybridizing with Rag1 probe

c) Map of the DNA samples loaded into the membrane

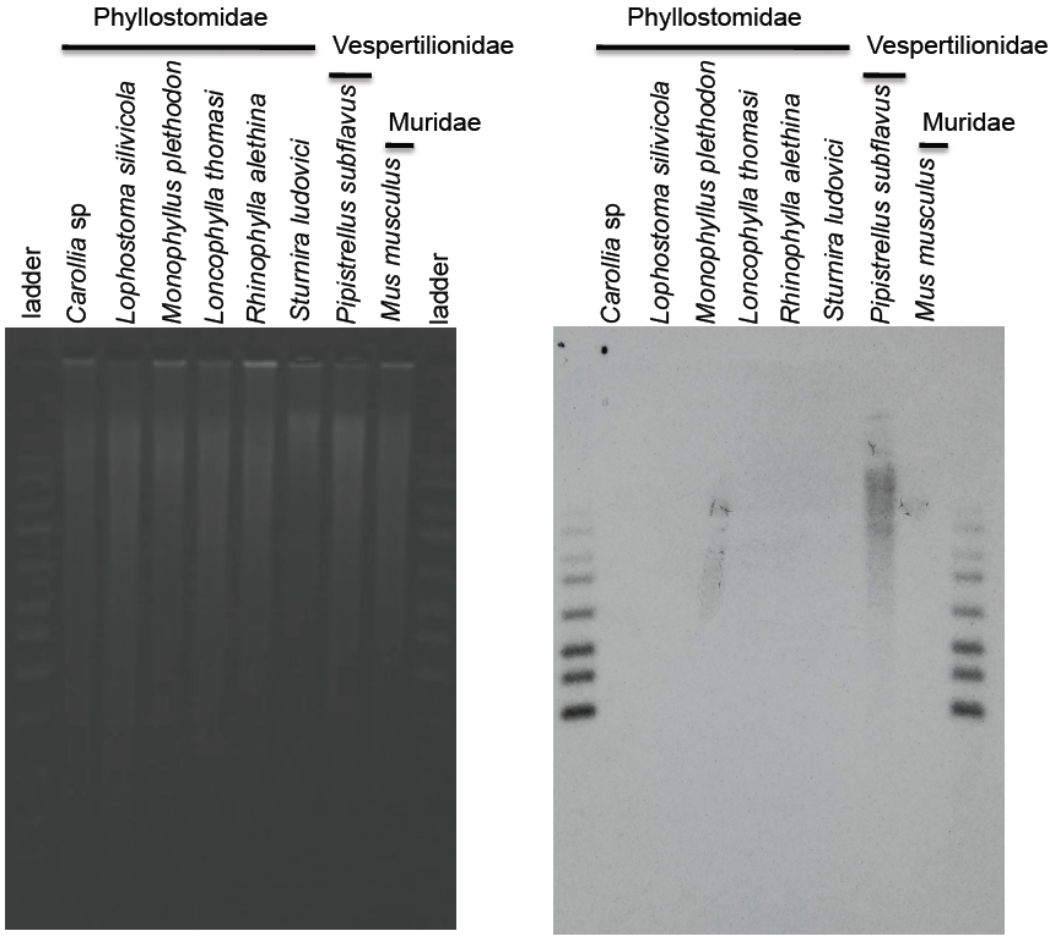
| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | *L. silivicolum* | *Myotis sp.* | *Miniopterus* | *Artibeus* | *Uroderma* | *S.lilium* |
| B | *B. cavernarum* | *Lophostoma* | *M.spasma* | *Loncophylla* | *S. ludovici* | *R.alethina* |
| C | *Vampyrum* | *Glischropus* | *Phyllostomus* | *Anoura geoffroyi* | *Mormoops* | *M. megalottis* |
| D | *L.thomasi* | *Balionycteris* | *Noctilio leporinus* | *Mus musculus* | *M. plethodon* | |
| E | 0.4 (199) | 1 (496) | 2.5 (1241) | | | |
| F | 0.4 (199) | 1 (496) | 2.5 (1241) | 30 (14894) | 10(4964) | 20 (7447) |

Genomic Slot Blot analysis. a) autoradiograph showing the hybridization of the membrane with the *Helibat* probe. b) autoradiograph showing the hybridization of the membrane with the Rag1 probe. c) Map of the DNA samples (1µg) loaded into the membrane and the combination of letters and numbers in the radiograph corresponds to the DNA samples loaded in the membrane. E and F lanes are loaded with the dilutions (ng) of the plasmid DNA containing the *Helitron* and Rag1 PCR products respectively and the numbers within the parenthesis indicates the copy number estimated by comparing the number of bat genomes (estimated genome size 2.3GB) and plasmid genomes (4.3kb) loaded onto the membrane. A, B, C, and D corresponds to the different bat species DNA and the mouse DNA (negative control) loaded. Details regarding the bat samples can be found in Supplemental Table 2.1. Genus identity of bat sample used in lane C2 was not confirmed through sequencing, but it was validated as a vesper bat.




a) 0.8% Agarose gel b) Autoradiograph

Genomic Southern blot analysis a) Genomic DNA samples of four different bat species and mouse (negative control) were restricted with BamH1 enzyme and were electrophoresed on a 0.8% agarose gel. b) Autoradiograph of nylon membrane blotted with DNA samples after hybridizing with the radio-labeled *Helibat* probe.




a) 0.8% Agarose gel

b) Autoradiograph

Genomic Southern blot analysis a) Genomic DNA samples of seven different bat species and mouse (negative control) were restricted with BamH1 enzyme and were electrophoresed on a 0.8% agarose gel. b) Autoradiograph of nylon membrane blotted with DNA samples after hybridizing with the radio-labeled *Helibat* probe.

APPENDIX F

EMPTY SITE CONFIRMATION FOR *HELITRONS* WITH GENE CAPTURES AND RETROGENES FOR NOVEL FAMILIES AND SUBFAMILIES OF *HELITRONS*

A)HelibatN12.1 (capture of serine threonine kinase)
AAPE01060724.1:898- 2818 AGACAGAGGCTGGGAGAGACTGGGATGGAGAGATGGAAACAAAATAGCAA/HB/TATTAAAATATTTCCTCTAATTAATTACCTTTTAATGTGCAT
AAPE01280493.1:14110-14011 AGACAGAGACTGGGAGAGACTGGGATGGAGAGATGGAAACAAAATAGCAA TATTAAAATATTTCCTCTAATTAATTCCCTTTTAATGTGCA
B) HelibatN1.26_
AAPE01095255.1|:2921-4706 GATTCTCATAATTTATAGTCTTCTTGTCTGTTCCCAAA-GGAATATTACAA//TGTATATATAATACTGAGAAAGCAAACTGAAAGAAACACTAAAA
AAWR02001015.1:c52938-52840 GAATCTCATAACTGACAGTGTTCTTATCTGCCCCCAAAATGAATAAAACAA--CGTATATAAAATACTGAGAAAGTAAATTGAAAGAAGCACTAAAA

C)HelibatN1.DD (DENNmadd domain, EIA)
AAPE01622385.1|:9371-10159 AGTTCTAAAGTGCTATGATTTCAGTCAAGTCCAACTTTTGATTAGTATTA//TTATATATGAGTATAGATTCTATAATTTGAAATGACATTTTATT
ABRP01001451.1| 2694-2787 AGTTCTAAAATGCTATGATTTCAGTCATGTCCAACTTTTGGTTAGCATCA--TTATATATAAATATAGATTGTATAATTTGAAATTATACTTTGTT

D)HelibatN1.2b (membrane associated ring finger)
AAPE01626805.1|:5490-6186 AACATTCCTGTTAATGAGGAATCACCTCCTTCCCCTTGGATTCAGTTTCA//TAGGACTATCAGGGAAGAGGCTGACCCATCCCTAGCCAAGCAGAA
ACTA01104283.1|:98937-99033AACATTCCCGTTAATGAGGAATTACTTCCTCCCCGTTGG-TTCAGTTTGA—-TAGGACTATAATGGGAGGGACT-CTCCATCTC-AGCCAAGGAGTA

E) Stat_1 (capture of protein inhibitor of stat1)
AAPE01478612.1|:2525-4045 CATGATCTCTAAAAGCAAGAGAGTATGTGGAGCAAAGACTCACACAGCAA//TTAGAGAATAAATACTGAATGGGTGATTAAATGGAGAAGCACATAT
ABRP01134837.1 :1958-2051 CAAGATCTCTAAAAGCAAGA--GTGTGTGGAGTAAAGACTCACATAGCAA--TTAGAAAGTAAAGACTGAATGAGTGGTTAAATGGAGAAGCGCACAT

F)53 GNAS complex locus
AAPE01312687.1|:470-1433 AATGTCAACATGGTTCAGATCACCCATAATTGTGCAGTTTAATTCTGTCA//TTTGAAAATAATA--GTGAC-ATTT-GTTCATTATGTAACTT
AEX02034779.1|:c82024-81933 AATGCCAATATGGTTTCGATCACTGATAATTGTGTGATTTAATTCTGTCA--TTTGAAATAATAATAGTAGCTATTTTGTTCATTTA-TAACTT

G) Helibat1.5q_N1
AAPE01051750.1|:6544-9264 TGAATGGCTTAAAATAGTGGGACACGGGGACTCTTAAAAAGAG-CAG-AGAA//TC-CTTTATAAATAGAGATAAAGCAGTTTCAGTTGTTTTGTAATA
ABRR01297778.1 :c394-292 ATGACTGAGAGTAGTTGTACACAGAGACTTTTAAAAAGAGAGAGGAGAA--TCACTTGATAAATAGACATAAAGCAGTTTCACTTGTTTTCTAATA

H)HelibatN1.26
AAPE02|cont2.8649:891-1717 TTTGTAAATTATATAATTGTCTAACCATTATGCTGTACACCTGAAACTAA//TATAAAATAATGTTGAATGTCAACTGTAATTGAAAAACAAA
AAPE02|cont2.50081:11607-11706 TTTGCAAATTATATAAATGTCTAATCACTATGCTATACACCTGAAACTAA--TATAAAATAATATTGAATGTCAACTGTAATTGAAAAATAAA

I) HelibatN1.5a3_bat1
AAPE01408579.1|:c2068-420 GTTAATGGCTAGATAAAACATATTCCATCCTCAGAATCTGGAGAAGCACA//TAGTTAAAAAATAAGATGTTCCCTGCTTATTTAACAAATAA
AAWR02010734.1 :c384022-383921 GTTAATGGCTAGATAAGACATACCCAGTTCTCAGGATCTAGAAAGCCCCA--TAGCT-AAAAATAAGACTCTCCATGCTTACTTACCAAATGA

J) HelibatN1.5t_N2
AAPE01108815.1|:3325-4726 AAACAACCCGAGTGTCCATTGACCAATAAATGGATAAGTAACATGTAGTA//TGTACATATAAGGGAATATTATTTAGCCTTAAAAAGGAAGGACATT
AAPE01429806.1|:6996-7095 AAACAGCTAGAGTGTCCATTGACCAATAAATGGATAAGTAACATGTAGTA--TGTAAATATAAGGGAATATTATTTGGCCTTAAAAAGGAAGGACATT

K)33_N2
AAPE01453066.1|:c6304-3573 GACTGGTTTCATAAATCCATCTGTACAATAAAATTAATATTATGTAGTCA//TTAAATATAATATTGATGTGTATGTAATGCATGTATTTGTG
gb|AAPE01051016.1|:c27040-26941 GACTGGTTTCATAAATCCATCTATACAATTAAATTAATATTATGTAGTCA--TTAAATATAATGTTGATGAGTATGTAATGCATGTATTTGTG

```
L)46_N2
AAPE01468138.1|:915-3533 GGCTTCCTCCGGAAGGACTCC-AGTCTAATTAGCATATAATGCTTTTATTA//TCTTATATAATAAAAGGCTAATATGCAAATTGTTCCCCTCGGAGTT
AAPE01225128.1|:743-842  GGCTTCATCTGGAAGGACGCCTGGTCTAATTAGCATATTATGCTTTTATTA--TTATAGATA-TAAAAGGCTAATATGCAAAGTGTCCACTCAGGAGTT

M) HelibatN2.5b_tandem
AAPE01596314.1|:c6525-4098  CAGCAACTTTGCAGACTGCCCTCTCGCATGCCAGGACCCCTATCTATGTA//TCTATGTATAATGCTAAGTGACTGGCCATCTGTGTGATGATTTTA
AAPE01091546.1|:c978-879    CAGCAACTTTGCATAGTGCCCTCCAGCACTCCGGGACCCCTATCTCTCTA--TATAAATAAAAGGCTAAGTGACTGGCCATAAGTCTGAGGATTTTA

N) 174
AAPE01460175.1|:205-2251 CAAAGTTTATTCAGCCAATCTACCTA----TGGATAGTTGGGTTGTTTTCCATA//TTTTGTAATAACAAACTAAGTTGCAATGAAATATTTTGTGCACA
ABRP01056938.1:1655-1755 CAAAGTTTATTCAGTCGATCT-CCTACATATGGACAGTTGGGTTGTTTTCAATA--TTTTGTAA--ACAAACTCAGTTGTAATGAATAACCTTATGAATA
Df) HelibatN2.12b
AAPE01216634.1|:4844-6324 CTGGACATGTATCTCCATTTTCAAAGAAGTAAGAGGATCTTGATTTCATA//TAAGATAT----------AATGATGTAAGTGGCTCTCAGTGGCACA
AAWR02002832.1 :2447-2554 CTGGACATGCATCTCTATTTTTAGAGAAGTAAGAGGATCTTGATCTCATA--TGACATATGGACATTATAAAATGATGCTAGTTGCTCTCAGTGGCACA

O)HelibatN2.15
AAPE01088121.1|:310-4766 TTGCAGGCCATGCCCCCCTCCAACCAGTACATGAATCCCTTGCTCTGGGCCTCAAA//TTTATGTATAAACTACTCAGGTTGAATGCTTTCAACACACAA
AAPE01572862.1|:c3077-2991                 CCTACCAGTGCATGAATTTGGTGCACTGGGCCTCTAG--TTTATGTATAAACTACTCAGGTTGAATGCTTTCAACACACAA

P) 214_GC
AAPE01523970.1|:4108-4764   AGAAAATATGA-ATTCAGTATTGGGCAGGGTTGGGTTTTAAATCCAGCTA//TAATAAATATTAGCTGACTATCTCCAGGCAAGTGGCCTTACAAC
ABRP01009538.1 :21601-21697 AGAAAACATGAGATTCAGTATTAGGCAGGGCTAGATTTTAATTCCAGTGA--TGCCAAATATTAGTGGAGTGTCTCCAGGCAAATGACCTTACAAC

Q) HelibatN1.4a_TAPT1GC
AAPE02009816.1:c2718-2109 GTATTGACAAGGACTTGCAGTTAAGATAATGCCGGTTTCACTTAAGAGTA//TTCTTGATAAAGCAGTAAAACATAGTTTTATTAACTATCAGCCCTG
ABRR01023120.1 :c5625-5722   TTGACAAGGACTTGCAGTTAAGGTAATGCCAGTATCACTTAGTAGTG--TTCTTAATGAAGGAGTAAAACATTAATTTAATTAAATCTCAACTCT

R) HelibatN1.30_N2
AAPE01623026.1|:c19531-16657 ATTCAACATATGAGGCATTTAGAAAATAATATTATATGTATATATAATATTACTA//TTTATAAATAATAAATAATAAAATCTTGGATTGAATGGT
ACTA01164548.1|:1857-1960     TTTAACATATGAAGCATTTAGTAAATTATTTTACATGTATATATAATAATATTA--TTTAGAAATAATAAATGATCAAATCTGGGATTGAATGGT

S)HelibatN1.17.1
AAPE01416846.1|:c3603-1632 TTCCTTAAAAGTGGAGGGCAAAACCATCCCAAAGTAACAGGAAAAATGCTA/HB_39/TTTAATATAAAAGCTTGAGCAAAGACTCTACTTCCTTTAA
ABRP|cont1.225182:c3411-3319 TTCCTTAAAAGTGGA---CAAAATCATCCCAAAGTAACGTGAATA---CTA     TTTAATATAAAAGCTTGAGAAAAGACTCTA--TCC-TTAA


T)HelibatN4.2
AAPE02|cont2.15726:1244-2254 AATAAGTCAGAGCAAGATAAATA-------ATCTC-TTATATGTGAAGTCTAAAACCA//TCATACTCATAAAAACAGAGAATAGATTGATGGTTA
ABRN|cont1.278420:c799-695   AATAAGTCAGAGAAAGACAAATACCATATGATCTCACTATATGTGGAATCTAAAAACG--TCAAACTCATAGAAACAGAGAATAGATTGATGGCTA


U)HelibatNa10 (587)
AAPE01635520.1|:3409-4067  TGCATCTTTTACTTTTGTTTCATCATGTA-GTTCTAAATGTGAGATGAAGAA//TTTAAAGATAAAAGCAAAAATACCTCATTCACCCTTTCTCCAA
WGS:ABRN|cont1.113361:7806-7898 CTTTTCCTTCTGTTTCATCGCGGATCGTCTAAATCTCAGAAAAAGAA--TTTAAAAATAAAAGCAAAAATATCTCATTTACCCTTTCTCTAC
```

```
V)HelibatN1.3e
AAPE01180129.1|:c4064-2933 AGCTAACTAAAATTTTTAAATGTACTTCAAAAATAAGATGTTATTCTACA//TAAAAATATATTAC--ATTAGTCTCACCAGCTATGTATGTAAATA
ABRP01302068.1 :c664-564   AGCTAATTGGAACTTTAAAATGAAATTCAAAA-TAAGATGTTATTTTACA--CAAAAATATATTACCTATTAGTCTCACCAGCTATGTTTGTAGAGA

W)HelibatN1.2a_a_N3
AAPE01237480.1|:2613-3382 TACCTGTAAAATGGGATGATAACAGTACTTACCTCCTAGGGCTATTGTAA//TCATTAAATAAGGTAACAGTTTGTGGAAGTGTCAACAGGGTAATAA
ABRP01093016.1 :1021-1118 TATCTGTAAAATGAGACAATAACAGTATTTACTTCCTAGAGCTGTTGTAA--TCATTAAATAAGATAATAGTTT-TGAAGCTTTCAACAGGTTAGTAA

X)194a
AAPE02|cont2.10245:34959-36242 TACAGCTTATGTTCAATA-ATTGTTATACAGATATACAATAGCTGACATAA//TTCTTCATAAAATTCCTTGATGTTTAGGAGAAGAAATTGGC
ABRP|cont1.240037:960-1060     TACAGCTTACTGTCCAGTGATTGTTGTACAGAAATACAATAGCTGACATAA--TTCTTCACAAAATTTCTTGATCTTTAGGAGGGTAAGTTGAC

Y)Helibat_Ribo
AAPE02|cont2.11003:37754-41255 ATGGTACTAAATGGAGACAGGAATAAAGCTAAAATTATAGGCAGAAATAA//TCCAAATATAATAAAGAAGTACATTATAGCATTATATAAGC
ABRP|cont1.82401:1479-1584     ATGGTACTAATTGAAGACAGTAATAAAGCTAAAATTATAGACCAAACTAA--CCCAAAGGTAATAAAGATGTGCATTATAGCATTATGTAAGC

Z) 40_N1
AAPE01229163.1|:1344-4029 TTTTCTCCTAAGTTGTCCAATGGTTGGCATATAATTGTTCATAGTATTTA//TATATGTATAAAAGCAAACTGTCCCCTTGGGAGTTTGCAAACTGTC
AAPE01336652.1|:c466-383   TTTCTGCTAGGTTGTCCATTTGTTGGCTTATAATTGTTCATAGTATCTA--TATATATA--AAAGCAAACTGTCCTCTTGGGAGTTTG

A1)HelibatNT_Ret
AAPE02033538.1:c26454-22131TCACCAATGGATTTCATGTAACCCACCTACATCCCTTTCGTTTGATTGCAA//TATATAAATATAGATGTAACCTGCCATTCTCCAGAGCATTATCTC
AAPE02|cont2.52990:455-556 TCACCAATGAATTTCATGTAACCCACCTACATCCCTTTCCTTTGATTGCAA--TGTATAAATACAGATGTAAGCCGCCATTCTCCAGAGCATTATCTC

A2)HelibatN1.24_N2
AAPE01027072.1|:c4958-1673 AAGAAAACCTAATGACACAGAAAGCACCCAAAAGTAGCATTATCCAATCA//TAAATTATAATTGCCTTTAACTCCTATTCCTAGTGCCCATTATTTC
ABRP01122519.1 :c6748-6843     AAATCTAATGAACTAGTAAGCACCATAACATAACATTTTTCAATCA--TAAGTGGTAATTGCCTTTTACTCCTACTCCTATTGCCCATTCTTTT


A3)Helibat1.3C
AAPE01581496.1 8125-4763 TGCTCCAGGAATAGGTTCCTGCCCTAGTATTTTTTTAATATAATAATAAGA//TTGTAATATAAAAATTGAAGTAACTTGTTTTTTAGTAGATGACACCT
AAPE01356402.1 330-429    TGCTCCAGGAATAGGCTCCTGCC-TAGTATTTTTTT-GTATAATAATAGGG--GTGTAATATAAAAATTGTAATAACTTGCTTTTTAGTAGATGACAACT
```

Empty site confirmation for *Helitrons* with gene captures and retrogenes.

```
AA) HelibatN9
AAPE01448368.1|:c2011-1647:ATATTAGGACTCTAGCCAAGGACAGTGACATTAAATTATAATCCAATATA/HB_N9/TTTTTCCATATAAAAACTTAACAGTTTGAAGAAGGTTTCAA
ABRP01088960.1|2561-2467       AGAACTCTAGCCAAGGACAGTGATGTTAAATCATAATTCAATATA/       TTTTTCATATAAAAATGT-AACAGTTTTAAGGAGGTTTCTA

AB)HelibatN8.1a
AAPE01517472.1: 173-571: ACGAATAACATTTCAGTCAATGATAAACTGTGTGTATGATGCTGGTCCCA// TAAGATTATAATGGAGCCAAATATTTCCTATTGCCTAGTGGCATTGT
ABRP01082900.1|:7498-7594   AATGACATTTCGATCAATGATAAATTGTGTGTATGATGGTGGTCCCA-- TAAGACTATAATGGAGCTGAAAAATTCCTATTGCCTAGTGACATAGT
```

```
AC) Helibatwith hat
AAPE02|cont2.56228:39968-40658 GATCTACCAGTCTTTAAGCTGAATGTCTTGAAGCTCTAAAATTTAGCAA//TACTATTATAAAATCACTTGTCAGAATCCCCCTAAATAGAAA
AAPE02|cont2.40411:c59763-59666 ATCTACTAGTCTTTAAGCTGAATGTCTTGAAGCTCTAAAAGTTATCAA--TATTATTATAAAATCGTGTGTCAGAATGCCCCTAAATAGAAA


AD)HelibatN1.2b
AAPE01626805.1|:5490-6186  AACATTCCTGTTAATGAGGAATCACCTCCTTCCCCTTGGATTCAGTTTCA//TAGGACTATCAGGGAAGAGGCTGACCCATCCCTAGCCAAGCAGAA
ACTA01104283.1|:98937-99033AACATTCCCGTTAATGAGGAATTACTTCCTCCCCGTTGG-TTCAGTTTGA—-TAGGACTATAATGGGAGGGACT-CTCCATCTC-AGCCAAGGAGTA

AE) helibat6.2
AAPE02|cont2.23442:c4670-4275   ATAT-CCTTTTTTTT-TTTTTTACTTTT--TTATTGAATTTATTGGAGGTGACA//TTGGTTAATAAAAATACAGATTTCAGGTGTACAATTC
AAPE02|cont2.24422:c12043-11941 ATATGCCTTTTCTTTATTTTTTATTTTTATTTATTGAATTTATTGG-GGTGACA--T-GGTTAATAAAATTATATAGGTTTCAGGTGTACAAT

AF)53
AAPE01312687.1|:470-1433     AATGTCAACATGGTTCAGATCACCCATAATTGTGCAGTTTAATTCTGTCA//TTTGAAAATAATA--GTGAC-ATTT-GTTCATTATGTAACTTT
AAEX02034779.1|:c82024-81933 AATGCCAATATGGTTTCGATCACTGATAATTGTGTGATTTAATTCTGTCA--TTTGAAATAATAATAGTAGCTATTTTGTTCATTTA-TAACTTT
AG)78
AAPE01191833.1|:2092-2502 ATCAAAATTTTAGCAG--TTTGAGTACTTGCTGGAAAATATAGTTGATATTATCA//TTTATAAATAATAGGAACACAGATCTTAAAAAGAGAAAGTA
ABRP01013946.1: 1032-1133    AAAATTTTAGTAAAATCTGAGTATTTGCTGGAAAATACAGTTAATATTATCA--TTTACAAAAAATAGGAATATAGATCTTAAAAAGAGAAAGTA

AH)215_a
AAPE01571732.1|:c1938-1556 CCATACGATCTAGCCATCTCACTTCTAGGTGTTTTCCCAAGAGAAATAAAAGCA//TATATTCATAACAA---GCCTGTGCAAATGTTCGAAGTAGC
AACN010328893.1|:c572-460  CCATATGATCCAATCATCCCACTTCTAGGTATTTTCCCAAGGGAAATGAAAGCA--TATGTTCATAAAAAATCTTCTATACAAATGTTCAAAGCAGC

AI)234
AAPE01079184.1|:6717-7794 GTGAACATTCCTTGTTTAACATTGCATAGCTTTGAAACATTTAATAATTA//TTGCTTAATAATAATGATTTACACTGCATACCATCTTTTAGAAATA
ABRP01033584.1 1452-1537  GTGAACGTTCCTTGTTTAATACTGCATACCTTTGAAATCTTTAATAATCA--TTGCTTAACAAT---GATTTATGCTGCACACCATCTTTT

AJ) 235
AAPE01631387.1|:1464-1953 TTTTAAAGCACTAATCTTTTTTC-TTTTTTATTGAATTTATTAAGGTGACA//TTGGTTAATAAAATTATATAGGTTTTAGGTGTACAATTCTATAAT
AAPE01521527.1|:1196-1296 TTTTCTAGCACCAAGTTTTATTATTTTTTTATTGAATTTATTGGGGTGACA--TTGGTTAACAAAATTATATAGGTTTCAGGTGTACAATTCTATAATA

AK)133
AAPE01356775.1|:c597-232GCAAAGAAATAATATTAAGGCATAAAGAGAAAA-----GAGG--------GATGGTAGAGATA//TTGAGAATAAAGGGATTGAGAAATTATGGATGAGCA
ACBE01469912.1|:c1469-1365      ATAATATTAAGGCATAAAGAGAAAATATTAGAGCAGCACAATGATGGTAGAGATA--TTGAGAACAAAGGGAATGAGAAGTTATGGATAACCA

AL)153
AAPE01078162.1|:c6094-5805 AGCATCAACTAATAACACTGCAG-TA----------ATTTGTAAATGCAGTGCAGT-GAACA//TGTATTCTAAAGAATCAGAACTCCTGCTAGCTT
ABRP01380284.1: c645-534   AGCATCAACTAATGACACATCAGCTAGCCCTCAGCCATTTGTAAATTCACACGCAGTAAACA—-CATATTCTAAAGAATCGGCATTCCTGCTAGCTT


AM)241_2
AAPE01300248.1|:c8243-7917 TGTTGTTTTTCATTAGATTATTTATTTTGATTTTTACATTCAATTGATGA//TAGATACATATGTATTGACATATTGTTGTTTCTATTTTTA
AAPE01027664.1|:902-988    --TTGTTTATTTTTAT-TTATTTATTTTGATTTTTAGATTCAATTGTTGA--TAGATATATATTTATTGCCATTTTATTGTTCATATTTTTA
```

```
AN)154
AAPE01626131.1|:13043-13272 TATACCTATCACTGCAAATGAGCAAAGTTTAGAATGTAGGAGGTAGCATA//TATCTTATGAATATGGAAAACACACAATACCCCTTTCTAAAAGT
ACBE01313708.1|:1052-1148   TATGCCTATCTCTGCAAATGAGCAAATTTTGGAATGTAGGAGGTAGAAAA--TAGGATATAAATATGAAAAGCAAACAATGCCCCTTTCTAAAAGT

AO)160
AAPE01375820.1|:c12494-12175 ACAAGCACTGGAGTTACAAGATGAAGATGACACAGCCCACCCCTGAACAA//TT---TACTATAAACTGAAAAAACAACAAACAAGCAGGCAATT
ABSL01077848.1|:c48258-48359 ACTAGCACTGGAATTACAAGATGAAGATGATACAGTCCACCCCTCAACAA—-TTGCATGCTACAACCTGAAAAACAAACAGACACACATGCAATT

AP)245
AAPE01400376.1|:c1100-1   TTACCACCCCATCAGTACAATGAAGTGAATTAGTTATGCAAATAA//TAATAATAATAAAACCTTGAGAGTATTTGCTTAGGAAGTTTTCAATA
AAPE01638397.1|:1037-1126   ACCACCCCATCAGTACAATGAAGTGAATTAGTTATGCAAATAA—TAATAATAATAAAACCTTGAGAGTATTTGCTTAGGAAGTTTTCAATA

AQ)250
AAPE01037770.1|:c2118-1149 CCACTGCAGGCCTTCACCACCCTATTGTCTGTGTACATAGGTAATGCATA//TATGCATATAAGATCTTTGTTTAATGTCTTTCTGCTCTCCACCCC
AAPE01178952.1|:1196-1295  CCACCCCAGGCCTTCGCCACCCTATTGTCTGTGTTCATAGGTAATGCATA--TATGCATATAAGATCTTTGTTTAATCTCTTCCTGTCCCCCATCCC

AR)184
AAPE01197819.1|:c3219-2892 CAGAAGGGTCCCTGGGACCCCTCAAAAGATTTGAAATTGCATGGGAAGCA//TTGTAAAATAAGTAATGAAAACTTCTAGGATACGTTGTGTAGGTA
ABRP01282681.1 :1419-1518  CAGAAGGGGCCCTAGAACATCCTAAAAGATTTAAAATTGCATGGGAAGCA--TTGTAAAATAAGTAATAAAAACTTCTAGGATATATTGTATGGGTG

AS)191
AAPE01289841.1|:4927-5706 GTTGACTTATCTCATCAGTGTTCTAAAATATTTGAGCGCCCCAGAAACA//TGTAGTCATAAAAGCTAAATTTTGATAACCAAGATCCTCTATTCCTC
ABRR01295802.1:3766-3861  GTTGACTTATCTCATCTGTGTTCCAAAATAT-CGAGTGCCCCAGAAACA--TTCAGCCATAAAAGCTAAATTATGAAAACCAAGATCCTCCACTTCTC

AT)192
AAPE02|cont2.36487:33093-33319 TTATAATAGTGCCTGGCACACACAATTAGCACTTAATAAATATTAGCCA//TTATTAATATTAACCACAATCCAGTTCATGAACAGGCTGCTT
ABRT|cont1.179974:c5320-5227   TTCTAATATTGCCTGACATACAAAATAAGCAATCAATAAGCATTAGCCA--TTATTAAGATGAATCACAATCCAGTTCATCAGCAGGGTGTTT


AU)238
AAPE02|cont2.2852:18538-19078 ATTGTTAAGTTACAAAATAATCACCAGGATCTAAAATACAGGATAAGAAA//TATAGTCAATAATATTGTAATAACTAGGTATGGTGCCAGAT
ABRP|cont1.242244:c536-440    ATTGGTAA-TTACAAAATAGTCACGGGGATGTAAAGTACAGCATAAGGAA--TATAGTCAATAATATTGTAATAACTATGTATGGTGCCAGGT


AV)HelibatN5.1b
AAPE02|cont2.59995:10504-10850ATAAAACTATAAAGCAGAGTCCTTGATTAGTTCCAGGGGAATT----CAGAGTA//TATGTGAA--TAAATGGCTCTACAGAGGAAGAAGGACG
ABRN|cont1.375530:583-676       AAAGCGATAAAGCAGAGTCCTTGATTAGTGCCAGGGGAATTAATTCAGAGCA--TATGTGAATGCATTAACTCTATTAGAGGAAGAAGGGCT


AW) 250
AAPE02|cont2.27402:45873-46842 CCACTGCAGGCCTTCACCACCCTATTGTCTGTGTACATAGGTAATGCATA//TATGCATATAAGATCTTTGTTTAATGTCTTTCTGCTCTCCA
AAPE02|cont2.47638:c3203-3106  CCACT-CAGGCCTTCACCACACTATTGTCTGTGTCCATAGGTAATGCATA--TATGCATATAAGATCTTTGTTTAATCTCTTCCTA-TCCCCA
```

```
AX)190
AAPE02|cont2.2887:106466-108280 ATTGAGATTTATTGAGATGACTATGACAATGAACCTCAGATCACCAACTA//TAAGGAATATAATTGATGAACAGCTTTAGTTACTGAGTGG
ABRP|cont1.254370:c8640-8550             TATTGAGAGGGCTACAAGAATGCATCTCAGACCTCCAACTA--TAAGGAATGTAATTGACTAAGGGCTTTAGCTGCTGTGTGG

AY)200

AAPE01243421.1|c2334-2106 CTACAATGCTTTTGCATATCCAGCTTGCCTTAGTCACTTGACTCTTGCAA//TCTTTTCAATAAGCATTTGGCTCCTAAGGAAATAGCCTTATATTTT
ABRP01224041.1:c1495-1400     AATGTTTTTGCGTATATTACTTGCCTTAGTCACTTGACTCTTAGAA--TCTTTTAAATAAGCATTTGGCTCCAAAAGAAACAGCCTTATATTTC

AZ)HelibatN3.2
AAPE01539716.1|:4664-5281  ATAAAACAATAAAACATAAAA-TTTAAGCTGTGA--GTCAATCAAATACAAAA//TTAGAAAATAAAAGAAAA-----GCATTAGCCTGGGATGTTG
ABRP01072693.1 :12862-12965ATAAAAAAATAAAATATAAAAATTTAAACTGTGAAAGTCAAACATATACAAAA--TAAAAAATAAAAATAAAATAAAAACATTAATTTGGGATGTTG

BA)HelibatNA27
AAPE01206816.1|:c1597-781      AAAAGACATTTTAACTAATTCTCTATGGCTTAATATTCATATCAGGGACA//TATAAGAATAAGTGTCTAATAGTTTAACTTCCATGAATGCA
AAWR02024209.1 :c121210-121110 AAAAGACATTTTAACCAATCATCTATAGCTTAACACACACATCAGGGATA--TCTAGGCATAAACGTCTAGTAGTTTAACATCCATGATTGCA

BB)236
AAPE02|cont2.52513:c985-530   ATCACATGATTTGTTCATGAAATTATCCTTCTGGGTGCACTGAATCAAAA//TGTTACATAAGATATTAATGAAAATTCTGGAATTT-ACTAAC
ABRP|cont1.244644:21571-21666    ACATGATTTGTTCATGAAATTATCTTTTGGGGTGCATGGAATCAAAA--TGTTAAATGAGATATTAATGAAAAATTGTAGAATTGACTAAC

BC)HelibatN1.5c
APE02|cont2.65881:c8274-6201 CTTTGTGATATTTAAATACAGTTTTCTTCAGATTCCAAGTAAGCAAACAA//TTCTTAAATAAAATGATATCGAATGTCTCATGAATTATTTCCAC
ABRP|cont1.152974:1178-1277  CTGTGTGATACTTAAATATAATTTTCTTCAGATTCCAAGTAAGCAAACAG--TTATTAAAGAAAATGTCTTTACATGTCTCATGAATGATTCCCAC

BD)198_
AAPE01603448.1|:c9533-9289 TGTTGACTTATTAACAATTTATCAAGAATAAGTACAACCATTGCAAAGTA//TTA---TATATATTTTTGAAAGATATGTTTGACCTAAGT
ABRP01053117.1 :c7756-7670   TTGACCTGTTAACAATTTATCAAGAGGATGTACAACTGTTGCAAAGTA--TTACTATACATACTTTTGAAAGATACATATGACCTAAGT
```

Empty site confirmation for novel families of *Helitrons*

```
CA) HelibatN1.30_a
AAPE01520718.1|:839-1993 ATTGATATTGTTTGAGAATCCATTGAATTATATGATCACCTAGTTTCCTAA//TAGAATTTAATAAAAACTATAGTTTTTCATCTTTCCACTGAAGATC
ACBE01034306.1|:2093-2193ATTGATATTGCTTGACAATCCATTGAATTATATGATCACCTAGTTTCCTGA--CAGAATTTAATTCAAACTATAGTTT-TCATCTTTCCACTACAGATT


CB)26
AAPE01390945.1|:105235-105686 TCAAGGACCCAGTTGCTGATCCCAACAAAAGGTCCAAAAAAGGTTGATTA//TCTTTACATAAAACACCAGAGGGAGATTTTGTTACACTTGAG
AAPE01535854.1|:7192-7291     TCAAGGACCCAGTTGCTGATCCCAACAAAAGGTCCAAGAAAGGTCGATTA--TCTTTACATAAAACACCAGCAGGAAATTTTGTTACACTTGAG

CD) 44
AAPE01240783.1|:c3233-2335   ATTGCAGTTAAAATTGATGCTACTTTTCTCTGCAGTTATATTTCACCCAA//TGTATAAATATTAATGTTAGTTTTTATTTCACTTTCCTTTGTA
AANU01206625.1|:c18853-18755 ATTACAGTTAAAATTGATGTTATTTTTTTCTTCATTTATATTTCATCCAA--TGTATGAAGATGAATGTTAGCTTTTATTTCTTTTTCCTTTGTA
```

```
CE) HelibatN2.12b
AAPE01216634.1|:4844-6324 CTGGACATGTATCTC      CAAAGAAGTAAGAGGATCTTGATTTCATA//TAAGATAT-----------AATGATGTAAGTGGCTCTCAGTGGCAC
AAWR02002832.1 :2447-2554 CTGGACATGCATCTCTATTTTTAGAGAAGTAAGAGGATCTTGATCTCATA--TGACATATGGACATTATAAAATGATGCTAGTTGCTCTCAGTGGCAC

CF)HelibatN1.2a_a_N3
AAPE01237480.1|:2613-3382 TACCTGTAAAATGGGATGATAACAGTACTTACCTCCTAGGGCTATTGTAA//TCATTAAATAAGGTAACAGTTTGTGGAAGTGTCAACAGGGTAATAA
ABRP01093016.1 :1021-1118 TATCTGTAAAATGAGACAATAACAGTATTTACTTCCTAGAGCTGTTGTAA--TCATTAAATAAGATAATAGTTT-TGAAGCTTTCAACAGGTTAGTAA


CG) HelibatN3.3c
AAPE01627569.1|:c2964-709   ATATATTACAAAACAGGAAAAGTCCCAAATCAATAATAATAACATGATCA//TTTATATAGAAAATC-AAAGGTATCTCCAAAACATTCCTAGAA
AAWR02002573.1 :27174-27266 ATACATTAGAAAATAGGACAAAATCTCAAGTCAAAATTATGACATGATTG--TCTGTGTAGAAAATCTAAAGGTATCTCCAAAACACTCCTAGAA


CH) HelibatN3.3d_N2
AAPE01389234.1|:c15536-13700 TATATTATACTAGTGACCCAGTGCACAAATTCCTACACATTGAAAAAAAA//TTAATTAGAAGAGATATTTTAATATTGCTATTCGCCCATTCTCT
AAPE01224067.1|:c1383-1284   TATTTTATACTAGTGACCCAGTGCACGGATTCGTGCACATTGAAAGGAAA--TTAATTAGAAGAAATATTTTAATATTGCTATTCGCCCTTTCTCT

CI) HelibatN3.3_N2
AAPE01095846.1|:c2741-630 ATGTTCATATATA-CCCACATTCAAAGACTGTTAAATCACGTTGTTTACCCA//TATCTTCAGAAAAAAATCGCTTCTGTCGTGGTAAACAACCTGCT
APE01621781.1|:2449-2550  ATGTTCAAATATATCCCACATTCAAAGGCTGTTAAATCGCGTTGTTTACCCA--TATCTTCAGAAAAAAATCACTTCTGTCGTGGTAAACAACCTGCT


CJ) Stat_1
AAPE01478612.1|:2525-4045 CATGATCTCTAAAAGCAAGAGAGTATGTGGAGCAAAGACTCACACAGCAA//TTAGAGAATAAATACTGAATGGGTGATTAAATGGAGAAGCACATAT
ABRP01134837.1 :1958-2051 CAAGATCTCTAAAAGCAAGA--GTGTGTGGAGTAAAGACTCACATAGCAA--TTAGAAAGTAAAGACTGAATGAGTGGTTAAATGGAGAAGCGCACAT

CK)HelibatN1.26
AAPE02|cont2.8649:891-1717     TTTGTAAATTATATAATTGTCTAACCATTATGCTGTACACCTGAAACTAA//TATAAAATAATGTTGAATGTCAACTGTAATTGAAAAACAAA
AAPE02|cont2.50081:11607-11706 TTTGCAAATTATATAAATGTCTAATCACTATGCTATACACCTGAAACTAA--TATAAAATAATATTGAATGTCAACTGTAATTGAAAAATAAA

CL)HelibatN1.24_N2
AAPE01027072.1|:c4958-1673 AAGAAAACCTAATGACACAGAAAGCACCCAAAAGTAGCATTATCCAATCA//TAAATTATAATTGCCTTTAACTCCTATTCCTAGTGCCCATTATTT
ABRP01122519.1 :c6748-6843     AAATCTAATGAACTAGTAAGCACCATAACATAACATTTTTCAATCA--TAAGTGGTAATTGCCTTTTACTCCTACTCCTATTGCCCATTCTTT


CM)210_N2
AAPE01048333.1|:9262-9706  GTTACTTAAAGACTAAGCAATAAAAGAGCAGGGTGATAGCCGTAAGAAGAA//TATGGCTATAAACCCATCTGAATTAATGTCATGACCATGTGAGA
AAFC03056107.1|:c8501-8401 GTTACTTAAAGACTAAGCAATGAAAGAGGAGGGTGATAAGCATAAGGAGAA--TGTGGCTATAAACACATCTGAATTAATCTCATGACCACGTGAGG
```

```
CN) HelibatN1.5s_a_bat1
AAPE01259266.1|:c1697-1045 AACCTCAAGAACTATCCTGTTTCTTAATAATGAAAAAAGGCTTGTTCACATCA//TCCTATTTAATAAAAGAGAAACATGTTAATTAGCCGTACCTCC
AAPE01587867.1|:2082-2184  AAACTCAAGAACTATCCAGTTTCTTAATAATGAAAAAAGGCTTGTTCACGTCA--TCCTATCTAATAAAAGAAAAACTTGGTAATTAGCCATATCTCC

CO) HelibatN1.13b_bat1
AAPE01573017.1|:1181-1482 GTTTAAATTATCCTAAAGATATAAACTGCTTATCTTATCCTATTTT-----------ACTA//TTTTTACATAACAGTTATCAGATGAATAGGGTAATA
ACTA01051671.1|:c4651-4543GTTTAAATTACCCTAGGGATATCTAC-GCTTATCTTATCCTATTTTGTACCCATGATACTA--TTTTTACCTAATAGTTGTCAGATTAATAGCATAATA


CP)HelibatN2.12a_bat1
AAPE01216634.1|:4844-6324 CTGGACATGTATCTCCATTTTCAAAGAAGTAAGAGGATCTTGATTTCATA//TAAGATAT-----------AATGATGTAAGTGGCTCTCAGTGGCACA
AAWR02002832.1 :2447-2554 CTGGACATGCATCTCTATTTTTAGAGAAGTAAGAGGATCTTGATCTCATA--TGACATATGGACATTATAAAATGATGCTAGTTGCTCTCAGTGGCACA

CQ) 40_N1
AAPE01229163.1|:1344-4029 TTTTCTCCTAAGTTGTCCAATGGTTGGCATATAATTGTTCATAGTATTTA//TATATGTATAAAAGCAAACTGTCCCCTTGGGAGTTTG
AAPE01336652.1|:c466-383   TTTCTGCTAGGTTGTCCATTTGTTGGCTTATAATTGTTCATAGTATCTA--TATATATA--AAAGCAAACTGTCCTCTTGGGAGTTTG

CR) HelibatN3.3a_bat1
AAPE01442638.1|:438-2266 AAGAAAAAACGTCATCCAAACATCAGGAAACAATGAGATATTTCCAGGAAA//TATTAAGATAAAATAGTAGGGAAAGTCTATCTGCATAATCAT
ACBE01221242.1|:6813-6906 AGAAAAATGCCATCCAAAGATCAGGAAACATGAACATTTTTAGAAAATA--TG--AAGATAAAGTACTAGGGAAAGCTTATCTG---TATCAT

Au) HelibatN2.9_bat1
AAPE01229701.1|:c2372-1871  GAGGCCAGGGATGATGCTAAACATTCCTAAACTACACAGGCCAGGCCACA//TAATGAATAATTATCTGGTCTAAAATGTCAATAGAGCCAATGTGA
AACN010124888.1|:c1641-1547 GAGGCCAAGATTGTTGCTAAATATTC-TACAATACACAGGAGAGGCCACA--TAAT----AATTATCTGGTCCAAAATGTCAATAGAGCCAGGGTGA

CS) HelibatN1.5a_bat1
AAPE01391598.1|:c10178-7831 CTTCCACTCTCTCTAAAAAATCAATGGAGAAATATCCT----TGA----TTAAAAAAA//TTTATATAAATAAAAGGCCCGTGGCCATCACACCAT
AAPE01306517.1|:c1281-1176  CTTCCACTCTCTCTAAAAAATCAATGGAGAAATATCCTCAAGTGAAGATTTAAAAAAA--TGAATATTAATAAAAGGCCCGTGGACGTCACACCAT


CT) HelibatN1.17_bat1
AAPE01416846.1|:c3603-1632  TTCCTTAAAAGTGGAGGGCAAAACCATCCCAAAGTAACAGGAAAAATGCTA//TTTAATATAAAAGCTTGAGCAAAGACTCTACTTCCTTTAAGTG
ABRP|cont1.225182:c3411-3319TTCCTTAAAAGTGGA---CAAAATCATCCCAAAGTAACGTGAATA---CTA--TTTAATATAAAAGCTTGAGAAAAGACTCTA--TCC-TTAAGTG

CU) Helibat1.4b_bat1
AAPE01355238.1|:c5688-3391 GGGATGGGGCAAGACCATAACCTTAGATGGCACTTAAGGTCGAATGATTA//TTATTCAAAATAAAAGGCAAGCAAATGAGAGGCAGTCAGATCCTAT
AAPE01525453.1|:1061-1160  GGGATGGGGCAAGACCATAACCTTAGATGGCACTTAAGGTCGAATTATTA--TTATTCAAAATAAAAGGCAAGTAAGTGAGAGGCAGTCAGATCCTAT

CV) Hb_GC1
AAPE01190778.1|:2378-3033 GCATGGTCAAGGAGTAAGAGAATGAGAGAACCAAAGCAGTTGTTATAATGA//TATGACATAAGAGCCATTAGATACGAAGAGTTCCCAAGCCCAACTC
AAWR02005348.1 :c394725-394632    AAGTATTAAAAAAATGAGAGAACCAAAGCTTTTGTTATAATGA--TGTGAAGAAAGACCCATTAGAACCTAAGAGTCCTAAAGCCCAGCTT

CW)191
AAPE01289841.1|:4927-5706 GTTGACTTATCTCATCAGTGTTCTAAAATATTTGAGCGCCCCAGAAACA//TGTAGTCATAAAAGCTAAATTTTGATAACCAAGATCCTCTATTCCTC
ABRR01295802.1:3766-3861  GTTGACTTATCTCATCTGTGTTCCAAAATAT-CGAGTGCCCCAGAAACA--TTCAGCCATAAAAGCTAAATTATGAAAACCAAGATCCTCCACTTCTC
```

```
CX)97_new
AAPE01137236.1|:900-1710 GGTATTGGATATGAGAATTCTTATTTATTTTGCTAAGGAAAAATATAGCA//TTGTGGATATAAAGGAAGTGGTTCTTATTTTTAAAAGACACCTG
ABRP01167384.1 : 787-878 GGTATTAGATAAGAGAGTTCTTATTAATTTTGCTAAGGGAAAAAT-GATA--TTGTGGTTTTAAAGGAAGTGGTCCTTATTTT-AAAAGACACATG

CY) 259
AAPE01040516.1|:c1696-1372 TAACAATGAAATTTAAATCAGTCCATGGAACCATTCCTAAAACAGATCAA//TAAAACAATAAGAAGCAATATCCTAAAGGACAGGGACATATATTTG
AAWR02029234.1:23757-23852      ATGAACTTTAAATCAATCCATAGAAATATTCCTAAGACAGATCAA--TAGAATACCAAGAAGCAATATCCTAAACAATAGGAAAATATATTTG
```

Empty site confirmation of novel *Helitron* subfamilies identified

# APPENDIX G

# CONCLUSIONS

We sought to address questions concerning the distribution of *Helitrons* among bats, the role of horizontal transfer (HT) in explaining their disparate distribution and their impact to *M. lucifugus*. Our analyses revealed that protein coding *Helitrons* seem to be limited to the vesper bats and implicated HT as a means to explain this distribution. We identify 37 families of *Helitrons* that represent two temporally independent invasions. The first invasion occurred in the ancestor of the vesper bats and the second in more recent time (~14 MYA). In addition, we show that 11.5% of the genome of *M. lucifugus* is composed of *Helitrons*, which is the largest percentage ever reported. We were able to leverage the availability of the extremely well annotated human genome and the slow rate of sequence evolution in mammals to show that many high copy number *Helitrons* have captured and amplified the promoter, exons, UTRs and introns of host genes and subjected them to rearrangement and shuffling. Previous genome wide reports of gene capture by *Helitrons* have only described captured protein-coding regions, which have not been further amplified to an appreciable frequency (e.g. Morgante *et al.* 2005, Lai *et al.* 2005, Yang and Bennetzen 2009). The identification of a few parental genes have provided for the first time, some insight into the mechanism by which *Helitrons* capture genes. Finally, numerous retrogenes were also identified in high-copy *Helitrons* and were occasionally shown to accompany the capture of promoter and other regulatory sequences from disparate loci. This novel discovery suggests that *Helitrons* might serve as generators of novel genes complete with the requisite regulatory apparatus and that this process may have led to the origin of novel genes in vesper bats.

Our future research plans include the investigation of the impact of *Helitron* mediated gene evolution on the transcriptome of vesper bats. An improved understanding of this process may shed light on the promise of *Helitrons* as gene therapy vectors.

REFERENCES


Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic Local Alignment Search Tool. Journal of Molecular Biology, 215, 403-410.

Baker, R. J., Olaf R.P., Bininda-Emonds., Mantilla-Meluk, H., Porter, C.A. and Van Den Bussche R.A.(in press). Molecular timescale of diversification of feeding strategy and morphology in new world leaf-nosed bats (Phyllostomidae): a phylogenetic perspective. In: "Evolutionary History of Bats: Fossils, Molecules and Morphology", edited by G. F. Gunnell and N. B. Simmons, Cambridge University Press, In the CUP series Cambridge Studies in Molecules and Morphology – New Evolutionary Paradigms

Bao, W.D., Jurka, M.G., Kapitonov, V.V. and Jurka, J. (2009) New Superfamilies of Eukaryotic DNA Transposons and Their Internal Divisions. Molecular Biology and Evolution, 26, 983-993.

Bartolome, C., Bello, X. and Maside, X. (2009) Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. Genome Biology, 10.

Barton, L.F., Runnels, H.A., Schell, T.D., Cho, Y.J., Gibbons, R., Tevethia, S.S., Deepe, G.S. and Monaco, J.J. (2004) Immune defects in 28-kDa proteasome activator gamma-deficient mice. Journal of Immunology, 172, 3948-3954.

Barton, R.M. and Worman, H.J. (1999) Prenylated prelamin A interacts with Narf, a novel nuclear protein. Journal of Biological Chemistry, 274, 30008-30018.

Biemont, C. and Vieira, C. (2006) Genetics - Junk DNA as an evolutionary force. Nature, 443, 521-524.

Bisson, I.A., Safi, K. and Holland, R.A. (2009) Evidence for Repeated Independent Evolution of Migration in the Largest Family of Bats. Plos One, 4.

Brambillasca, F., Mosna, C., Colombo, M., Rivolta, A., Caslini, C., Minuzzo, M., Giudici, G., Mizzi, L., Biondi, A. and Privitera, E. (1999) Identification of a novel molecular partner of the E2A gene in childhood leukemia. Leukemia, 13, 369-375.

Brambillasca, F., Mosna, G., Ballabio, E., Biondi, A., Boulukos, K.E. and Privitera, E. (2001) Promoter analysis of TFPT (FB1), a molecular partner of TCF3 (E2A) in childhood acute lymphoblastic leukemia. Biochemical and Biophysical Research Communications, 288, 1250-1257.

Britten, R. (2006) Transposable elements have contributed to thousands of human proteins. Proceedings of the National Academy of Sciences of the United States of America, 103, 1798-1803.

Brunner, S., Pea, G. and Rafalski, A. (2005) Origins, genetic organization and transcription of a family of non-autonomous *Helitron* elements in maize. Plant Journal, 43, 799-810.

Calisher, C.H., Childs, J.E., Field, H.E., Holmes, K.V. and Schountz, T. (2006) Bats: Important reservoir hosts of emerging viruses. Clinical Microbiology Reviews, 19, 531-+.

Casse, N., Bui, Q.T., Nicolas, V., Renault, S., Bigot, Y. and Laulier, M. (2006) Species sympatry and horizontal transfers of Mariner transposons in marine crustacean genomes. Molecular Phylogenetics and Evolution, 40, 609-619.

Chen, J.M., Stenson, P.D., Cooper, D.N. and Ferec, C. (2005) A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Human Genetics, 117, 411-427.

Choi, J.D., Hoshino, A., Park, K.I., Park, I.S. and Iida, S. (2007) Spontaneous mutations caused by a *Helitron* transposon, Hel-It1, in morning glory, *Ipomoea tricolor*. Plant Journal, 49, 924-934.

Craig, N. L., 2002. Mobile DNA: an introduction, in: Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. (Eds.), Mobile DNA II. American Society of Microbiology Press., Washington D.C. pp 3-11.

Cultrone, A., Dominguez, Y.R., Drevet, C., Scazzocchio, C. and Fernandez-Martin, R. (2007) The tightly regulated promoter of the xanA gene of *Aspergillus nidulans* is included in a *Helitron*. Molecular Microbiology, 63, 1577-1587.

de Boer, J.G., Yazawa, R., Davidson, W.S. and Koop, B.F. (2007) Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. Bmc Genomics, 8.

Deininger, P.L. and Batzer, M.A. (1999) Alu repeats and human disease. Molecular Genetics and Metabolism, 67, 183-193.

Diao, X.M., Freeling, M. and Lisch, D. (2006) Horizontal transfer of a plant transposon. Plos Biology, 4, 119-128.

Du, C., Fefelova, N., Caronna, J., He, L.M. and Dooner, H.K. (2009) The polychromatic *Helitron* landscape of the maize genome. Proceedings of the National Academy of Sciences of the United States of America, 106, 19916-19921.

Ejima, Y. and Yang, L.C. (2003) Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. Human Molecular Genetics, 12, 1321-1328.

Feschotte, C. (2008) Opinion - Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics, 9, 397-405.

Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. Annual Review of Genetics, 41, 331-368.

Feschotte, C. and Pritham, E.J. (2009) A cornucopia of *Helitron*s shapes the maize genome. Proceedings of the National Academy of Sciences of the United States of America, 106, 19747-19748.

Feschotte, C. and Wessler, S.R. (2001) Treasures in the attic: Rolling circle transposons discovered in eukaryotic genomes. Proceedings of the National Academy of Sciences of the United States of America, 98, 8923-8924.

Feschotte, C., Keswani, U., Ranganathan, N., Guibotsy, M.L. and Levine, D. (2009) Exploring Repetitive DNA Landscapes Using REPCLASS, a Tool That Automates the Classification of Transposable Elements in Eukaryotic Genomes. Genome Biology and Evolution, 1, 205-220.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L. and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics, 151, 1531-1545.

Franchini, C., Fontana, F., Minuzzo, M., Babbio, F. and Privitera, E. (2006) Apoptosis promoted by up-regulation of TFPT (TCF3 fusion partner) appears p53 independent, cell type restricted and cell density influenced. Apoptosis, 11, 2217-2224.

Galanty, Y., Belotserkovskaya, R., Coates, J., Polo, S., Miller, K.M. and Jackson, S.P. (2009) Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. Nature, 462, 935-U132.

Gan, Y.H., Taira, E., Irie, Y., Fujimoto, T. and Miki, N. (2003) Arrest of cell cycle by Amida which is phosphorylated by Cdc2 kinase. Molecular and Cellular Biochemistry, 246, 179-185.

Gergely, F., Draviam, V.M. and Raff, J.W. (2003) The ch-TOG/XMAP215 protein is essential for spindle pole organization in human somatic cells. Genes & Development, 17, 336-341.

Gilbert, C., Schaack, S., Pace, J.K., Brindley, P.J. and Feschotte, C. (2010) A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature, 464, 1347-U1344.

Gray, Y.H.M. (2000) It takes two transposons to tango - transposable-element-mediated chromosomal rearrangements. Trends in Genetics, 16, 461-468.

Gupta, S., Gallavotti, A., Stryker, G.A., Schmidt, R.J. and Lal, S.K. (2005) A novel class of *Helitron*-related transposable elements in maize contain portions of multiple pseudogenes. Plant Molecular Biology, 57, 115-127.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H.Y., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T. et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature, 461, 393-398.

Hao, Z.L., Stoler, M.H., Sen, B., Shore, A., Westbrook, A., Flickinger, C.J., Herr, J.C. and Coonrod, S.A. (2002) TACC3 expression and localization in the murine egg and ovary. Molecular Reproduction and Development, 63, 291-299.

Hart, D.L., Lohe, A.R. and Lozovskaya, E.R. (1997) Modern thoughts on an ancyent marinere: Function, evolution, regulation. Annual Review of Genetics, 31, 337-358.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A. and Wendel, J.F. (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Research, 16, 1252-1261.

Hecht, M.M., Nitz, N., Araujo, P.F., Sousa, A.O., Rosa, A.D., Gomes, D.A., Leonardecz, E. and Teixeira, A.R.L. (2010) Inheritance of DNA Transferred from American Trypanosomes to Human Hosts. Plos One, 5.

Hoffmann, F.G. and Baker, R.J. (2001) Systematics of bats of the genus Glossophaga (Chiroptera : Phyllostomidae) and phylogeography in G-Soricina based on the cytochrome-b gene. Journal of Mammalogy, 82, 1092-1101.

Hoffmann, F.G., Hoofer, S.R. and Baker, R.J. (2008) Molecular dating of the diversification of Phyllostominae bats based on nuclear and mitochondrial DNA sequences. Molecular Phylogenetics and Evolution, 49, 653-658.

Hollister, J.D. and Gaut, B.S. (2007) Population and evolutionary dynamics of *Helitron* transposable elements in *Arabidopsis thaliana*. Molecular Biology and Evolution, 24, 2515-2524.

Hoshino, A., Choi, J.D., Park, K.I., Park, I.S. and Da, S.I. (2007) Spontaneous mutations caused by a *Helitron* transposon, *Hel-It1*, in *Ipomoea tricolor*. Plant and Cell Physiology, 48, S152-S152.

Hotopp, J.C.D., Clark, M.E., Oliveira, D., Foster, J.M., Fischer, P., Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S.L. et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science, 317, 1753-1756.

Houck, M.A., Clark, J.B., Peterson, K.R. and Kidwell, M.G. (1991) Possible horizontal transfer of drosophila genes by the mite *Proctolaelaps regalis*. Science, 253, 1125-1129.

Houlden, H., Johnson, J., Gardner-Thorpe, C., Lashley, T., Hernandez, D., Worth, P., Singleton, A.B., Hilton, D.A., Holton, J., Revesz, T. et al. (2007) Mutations in TTBK2, encoding a kinase implicated in tau phosphorylation, segregate with spinocerebellar ataxia type 11. Nature Genetics, 39, 1434-1436.

Huang, Y., Niu, B.F., Gao, Y., Fu, L.M. and Li, W.Z. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics, 26, 680-682.

Irie, Y., Yamagata, K., Gan, Y.H., Miyamoto, K., Do, E., Kuo, C.H., Taira, E. and Miki, N. (2000) Molecular cloning and characterization of amida, a novel protein which interacts with a neuron-specific immediate early gene product arc, contains novel nuclear localization signals, and causes cell death in cultured cells. Journal of Biological Chemistry, 275, 2647-2653.

Jiang, N., Bao, Z.R., Zhang, X.Y., Eddy, S.R. and Wessler, S.R. (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature, 431, 569-573.

Jones, M.C., Fusi, L., Higham, J.H., Abdel-Hafiz, H., Horwitz, K.B., Lam, E.W.F. and Brosens, J.J. (2006) Regulation of the SUMO pathway sensitizes differentiating human endometrial stromal cells to progesterone. Proceedings of the National Academy of Sciences of the United States of America, 103, 16272-16277.

Kalka, M.B., Smith, A.R. and Kalko, E.K.V. (2008) Bats limit arthropods and herbivory in a tropical forest. Science, 320, 71-71.

Kapitonov, V.V. and Jurka, J. (2001) Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America, 98, 8714-8719.

Kapitonov, V.V. and Jurka, J. (2007) *Helitron*s on a roll: eukaryotic rolling-circle transposons. Trends in Genetics, 23, 521-529.

Kazazian, H.H. (2004) Mobile elements: Drivers of genome evolution. Science, 303, 1626-1632.

Keeling, P.J. and Palmer, J.D. (2008) Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics, 9, 605-618.

Kidwell, M.G. (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica, 115, 49-63.

Kidwell, M.G. (1992) Horizontal transfer of p-elements and other short inverted repeat transposons. Genetica, 86, 275-286.

Lai, J.S., Li, Y.B., Messing, J. and Dooner, H.K. (2005) Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. Proceedings of the National Academy of Sciences of the United States of America, 102, 9068-9073.

Lal, S., Oetjens, M. and Hannah, L.C. (2009) *Helitron*s: Enigmatic abductors and mobilizers of host genome sequences. Plant Science, 176, 181-186.

Lal, S.K. and Hannah, L.C. (2005) *Helitron*s contribute to the lack of gene colinearity observed in modern maize inbreds. Proceedings of the National Academy of Sciences of the United States of America, 102, 9993-9994.

Lal, S.K. and Hannah, L.C. (2005) Plant genomes - Massive changes of the maize genome are caused by *Helitron*s. Heredity, 95, 421-422.

Lal, S.K., Giroux, M.J., Brendel, V., Vallejos, C.E. and Hannah, L.C. (2003) The maize genome contains a *Helitron* insertion. Plant Cell, 15, 381-391.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome. Nature, 409, 860-921.

Langdon, T., Thomas, A., Huang, L., Farrar, K., King, J. and Armstead, I. (2009) Fragments of the key flowering gene GIGANTEA are associated with *Helitron*-type sequences in the Pooideae grass Lolium perenne. Bmc Plant Biology, 9.

Li, L.L., Victoria, J.G., Wang, C.L., Jones, M., Fellers, G.M., Kunz, T.H. and Delwart, E. (2010) Bat Guano Virome: Predominance of Dietary Viruses from Insects and Plants plus Novel Mammalian Viruses. Journal of Virology, 84, 6955-6965.

Lim, J.K. and Simmons, M.J. (1994) Gross chromosome rearrangements mediated by transposable elements in drosophila-melanogaster. Bioessays, 16, 269-275.

Liu, B., Liao, J.Y., Rao, X.P., Kushner, S.A., Chung, C.D., Chang, D.D. and Shuai, K. (1998) Inhibition of Stat1-mediated gene activation by PIAS1. Proceedings of the National Academy of Sciences of the United States of America, 95, 10626-10631.

Long, M., Betran, E., Thornton, K. and Wang, W. (2003) The origin of new genes: Glimpses from the young and old. Nature Reviews Genetics, 4, 865-875.

Loreto, E.L.S., Carareto, C.M.A. and Capy, P. (2008) Revisiting horizontal transfer of transposable elements in Drosophila. Heredity, 100, 545-554.

Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A. and Kaessmann, H. (2005) Emergence of young human genes after a burst of retroposition in primates. Plos Biology, 3, 1970-1979.

Marquez, C.P. and Pritham, E.J. (2010) Phantom, a New Subclass of Mutator DNA Transposons Found in Insect Viruses and Widely Distributed in Animals. Genetics, 185, 1507-U1582.

Maystadt, I., Rezsohazy, R., Barkats, M., Duque, S., Vannuffel, P., Remacle, S., Lambert, B., Najimi, M., Sokal, E., Munnich, A. et al. (2007) The nuclear factor kappa B-activator gene PLEKHG5 is mutated in a form of autosomal recessive lower motor neuron disease with childhood onset. American Journal of Human Genetics, 81, 67-76.

Maystadt, I., Zarhrate, M., Leclair-Richard, D., Estournet, B., Barois, A., Renault, F., Routon, M.C., Durand, M.C., Lefebvre, S., Munnich, A. et al. (2006) A gene for an autosomal recessive lower motor neuron disease with childhood onset maps to 1p36. Neurology, 67, 120-124.

Miller-Butterworth, C.M., Murphy, W.J., O'Brien, S.J., Jacobs, D.S., Springer, M.S. and Teeling, E.C. (2007) A family matter: Conclusive resolution of the taxonomic position of the long-fingered bats, Miniopterus. Molecular Biology and Evolution, 24, 1553-1561.

Moran, J.V., DeBerardinis, R.J. and Kazazian, H.H. (1999) Exon shuffling by L1 retrotransposition. Science, 283, 1530-1534.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A. and Rafalski, A. (2005) Gene duplication and exon shuffling by *Helitron*-like transposons generate intraspecies diversity in maize. Nature Genetics, 37, 997-1002.

Nekrutenko, A. and Li, W.H.S. (2001) Transposable elements are found in a large number of human protein-coding genes. Trends in Genetics, 17, 619-621.

Neuweiler, G., (2000). The biology of Bats. Oxford University Press, New York.

Novick, P., Smith, J., Ray, D. and Boissinot, S. (2010) Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. Gene, 449, 85-94.

Nowack, R. M.1994. Walkers Bats of the World. The Johns Hopkins University Press, Baltimore and London.

Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. Nature, 405, 299-304.

Oliver, K.R. and Greene, W.K. (2009) Transposable elements: powerful facilitators of evolution. Bioessays, 31, 703-714.

Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA - The ultimate parasite. Nature, 284, 604-607.

Pace, J.K. and Feschotte, C. (2007) The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. Genome Research, 17, 422-432.

Pace, J.K., Gilbert, C., Clark, M.S. and Feschotte, C. (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proceedings of the National Academy of Sciences of the United States of America, 105, 17023-17028.

Pagan, H.J.T., Smith, J.D., Hubley, R.M. and Ray, D.A. (2010) PiggyBac-ing on a Primate Genome: Novel Elements, Recent Activity and Horizontal Transfer. Genome Biology and Evolution, 2, 293-303.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D.S., Jackson, S., Wing, R.A. et al. (2006) Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Research, 16, 1262-1269.

Piriyapongsa, J., Marino-Ramirez, L. and Jordan, I.K. (2007) Origin and evolution of human microRNAs from transposable elements. Genetics, 176, 1323-1337.

Poulter, R.T.M., Goodwin, T.J.D. and Butler, M.I. (2003) Vertebrate helentrons and other novel *Helitron*s. Gene, 313, 201-212.

Price, A.L., Jones, N.C. and Pevzner, P.A. (2005) De novo identification of repeat families in large genomes. Bioinformatics, 21, I351-I358.

Pritham, E.J. (2009) Transposable Elements and Factors Influencing their Success in Eukaryotes. Journal of Heredity, 100, 648-655.

Pritham, E.J. and Feschotte, C. (2007) Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. Proceedings of the National Academy of Sciences of the United States of America, 104, 1895-1900.

Ray, D.A., Feschotte, C., Pagan, H.J.T., Smith, J.D., Pritham, E.J., Arensburger, P., Atkinson, P.W. and Craig, N.L. (2008) Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. Genome Research, 18, 717-728.

Ray, D.A., Pagan, H.J.T., Thompson, M.L. and Stevens, R.D. (2007) Bats with hATs: Evidence for recent DNA transposon activity in genus Myotis. Molecular Biology and Evolution, 24, 632-639.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y. et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science, 319, 64-69.

Roulin, A., Piegu, B., Fortune, P.M., Sabot, F., D'Hont, A., Manicacci, D. and Panaud, O. (2009) Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae. Bmc Evolutionary Biology, 9.

Schaack, S., Gilbert, C. and Feschotte, C. (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends in Ecology & Evolution, 25, 537-546.

Silva, J.C. and Kidwell, M.G. (2000) Horizontal transfer and selection in the evolution of P elements. Molecular Biology and Evolution, 17, 1542-1557.

Silva, J.C., Loreto, E.L. and Clark, J.B. (2004) Factors that affect the horizontal transfer of transposable elements. Current Issues in Molecular Biology, 6, 57-71.

Sorek, R., Ast, G. and Graur, D. (2002) Alu-containing exons are alternatively spliced. Genome Research, 12, 1060-1067.

Stadelmann, B., Lin, L.K., Kunz, T.H. and Ruedi, M. (2007) Molecular phylogeny of New World Myotis (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. Molecular Phylogenetics and Evolution, 43, 32-48.

Sweredoski, M., DeRose-Wilson, L. and Gaut, B.S. (2008) A comparative computational analysis of nonautonomous *Helitron* elements between maize and rice. Bmc Genomics, 9.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution, 24, 1596-1599.

Tan, I., Ng, C.H., Lim, L. and Leung, T. (2001) Phosphorylation of a novel myosin binding subunit of protein phosphatase 1 reveals a conserved mechanism in the regulation of actin cytoskeleton. Journal of Biological Chemistry, 276, 21209-21216.

Teeling, E.C., Springer, M.S., Madsen, O., Bates, P., O'Brien, S.J. and Murphy, W.J. (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. Science, 307, 580-584.

Teeling, E.C., Springer, M.S., Madsen, O., Bates, P., O'Brien, S.J. and Murphy, W.J. (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. Science, 307, 580-584.

Tempel, S., Nicolas, J., El Amrani, A. and Couee, I. (2007) Model-based identification of *Helitron*s results in a new classification of their families in Arabidopsis thaliana. Gene, 403, 18-28.

Thomas, J., Schaack, S. and Pritham, E.J. (2010) Pervasive Horizontal Transfer of Rolling-Circle Transposons among Animals. Genome Biology  Evolution, 2, 656-664.

Toleman, M.A., Bennett, P.M. and Walsh, T.R. (2006) ISCR elements: Novel gene-capturing systems of the 21st century? Microbiology and Molecular Biology Reviews, 70, 296-+.

Volff, J.N. (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays, 28, 913-922.

Vonhof, M.J., Barber, D., Fenton, M.B. and Strobeck, C. (2006) A tale of two siblings: multiple paternity in big brown bats (Eptesicus fuscus) demonstrated using microsatellite markers. Molecular Ecology, 15, 241-247.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature, 420, 520-562.

Williams-Guillen, K., Perfecto, I. and Vandermeer, J. (2008) Bats limit insects in a neotropical agroforestry system. Science, 320, 70-70.

Xing, J., Wang, H., Belancio, V.P., Cordaux, R., Deininger, P.L. and Batzer, M.A. (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. Proceedings of the National Academy of Sciences of the United States of America, 103, 17608-17613.

Xu, J.H. and Messing, J. (2006) Maize haplotype with a *Helitron*-amplified cytidine deaminase gene copy. Bmc Genetics, 7.

Yang, L.X. and Bennetzen, J.L. (2009a) Structure-based discovery and description of plant and animal *Helitron*s. Proceedings of the National Academy of Sciences of the United States of America, 106, 12832-12837.

Yang, L.X. and Bennetzen, J.L. (2009b) Distribution, diversity, evolution, and survival of *Helitron*s in the maize genome. Proceedings of the National Academy of Sciences of the United States of America, 106, 19922-19927.

Yi, S.J., Ellsworth, D.L. and Li, W.H. (2002) Slow molecular clocks in Old World monkeys, apes, and humans. Molecular Biology and Evolution, 19, 2191-2198.

Yoshiyama, M., Tu, Z., Kainoh, Y., Honda, H., Shono, T. and Kimura, K. (2001) Possible horizontal transfer of a transposable element from host to parasitoid. Molecular Biology and Evolution, 18, 1952-1958.

Zeh, D.W., Zeh, J.A. and Ishida, Y. (2009) Transposable elements and an epigenetic basis for punctuated equilibria. Bioessays, 31, 715-726.

Zhang, J.Z. (2003) Evolution by gene duplication: an update. Trends in Ecology & Evolution, 18, 292-298.

Zheng, X.J., Xu, C., Di Lorenzo, A., Kleaveland, B., Zou, Z.Y., Seiler, C., Chen, M., Cheng, L., Xiao, J.P., He, J. et al. (2010) CCM3 signaling through sterile 20-like kinases plays an essential role during zebrafish cardiovascular development and cerebral cavernous malformations. Journal of Clinical Investigation, 120, 2795-2804.

Zhou, Q., Froschauer, A., Schultheis, C., Schmidt, C., Bienert, G.P., Wenning, M., Dettai, A. and Volff, J.-N. (2006) *Helitron* transposons on the sex chromosomes of the platyfish Xiphophorus maculatus and their evolution in animal genomes. Zebrafish, 3, 39-52.

Zhou, T.H., Ling, K., Guo, J., Zhou, H., Wu, Y.L., Jing, Q., Ma, L. and Pei, G. (2000) Identification of a human brain-specific isoform of mammalian STE20-like kinase 3 that is regulated by cAMP-dependent protein kinase. Journal of Biological Chemistry, 275, 2513-2519.

## BIOGRAPHICAL INFORMATION

Jainy Thomas was born in Kerala, India to C.V.Thomas and Mary Thomas. She completed her Bachelor's in Agriculture in Kerala Agriculture University, India in 2003. Through a highly competitive exam, she gained admission in the prestigious Indian Agricultural Research Institute, New Delhi, India and was awarded Junior Research Fellowship. She worked on the impact of cyst nematode (*Heterodera avenae*) on the physiological and quality parameters of wheat and completed her masters in Nematology in 2006. She is married to Tharun J. Puthenkandom in 2007. To pursue her dream of becoming a scientist, she joined for the doctoral program in the University of Texas at Arlington. She worked on Mobile DNA and has presented her research in several national and international meetings and has received many awards including travel grants. She is planning to do her post-doctoral research and aim for becoming a scientist.