CrossMark

Mammalian transposable elements and their impacts on genome evolution

Roy N. Platt II · Michael W. Vandewege · David A. Ray

Abstract Transposable elements (TEs) are genetic elements with the ability to mobilize and replicate themselves in a genome. Mammalian genomes are dominated by TEs, which can reach copy numbers in the hundreds of thousands. As a result, TEs have had significant impacts on mammalian evolution. Here we summarize the current understanding of TE content in mammal genomes and find that, with a few exceptions, most fall within a predictable range of observations. First, one third to one half of the genome is derived from TEs. Second, most mammalian genomes are dominated by LINE and SINE retrotransposons, more limited LTR retrotransposons, and minimal DNA transposon accumulation. Third, most mammal genome contains at least one family of actively accumulating retrotransposon. Finally, horizontal transfer of TEs among lineages is rare. TE exaptation events are being recognized with increasing frequency. Despite these beneficial aspects of TE content and activity, the majority of TE insertions are neutral or deleterious. To limit the deleterious effects of TE proliferation, the genome has evolved several defense mechanisms that act at the epigenetic, transcriptional, and post-transcriptional levels. The interaction between TEs and these defense mechanisms has led to an evolutionary arms race where TEs are suppressed, evolve to escape suppression, then are suppressed again as the defense mechanisms undergo compensatory change. The result is complex and constantly evolving interactions between TEs and host genomes.

Abbreviations

TE	Transposable element
ERV	Endogenous retrovirus
LTR	Long terminal repeat
LINE	Long interspersed element
SINE	Short interspersed element
Kb	Kilobases
Mb	Megabases
Bp	Base pairs
ORF	Open reading frame
L1	LINE1
L2	LINE2
RTE	RNA transport element
MY	Million years
MYA	Million years ago
piRNA	PIWI-interacting RNA
KRAB-ZFPs	KRAB zinc finger proteins
APOBEC	Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like

R. N. Platt II, (✉) · M. W. Vandewege · D. A. Ray
Department of Biological Sciences, Texas Tech University,
Lubbock, TX, USA
e-mail: neal.platt@gmail.com

Introduction

Genome evolution is a highly dynamic process where large-scale genomic change can occur through a range of events including whole genome duplications, inversions, segmental duplications or deletions, and transposable element (TE) insertions and excisions (Cheng et al. 2005; Franke et al. 2017; Marques-Bonet et al. 2009; Ohno 1970). TEs are selfish genetic elements with the ability to mobilize within a genome. During the mobilization process, new copies of the TE can be created directly (Ostertag and Kazazian Jr. 2001), or indirectly, depending on the TE type or position of the TE relative to a DNA replication fork (Chen et al. 1992). TEs often reach high copy number over short evolutionary periods because of their replicative nature and continuous accumulation.

In most cases, TE insertions have no identified function (Biémont 2010) but examples of exapted TE insertions are becoming increasingly common (reviewed in Warren et al. 2015). Function has been ascribed to individual insertions (Mi et al. 2000), entire TE families (Bourque et al. 2008), and TEs in general (Cowley and Oakey 2013). Despite the potential advantages TEs provide, many insertions are neutral or deleterious, potentially resulting in a disease-state, or even a lethal allele. As a result, several genome defense mechanisms have evolved to limit TE activity. The goal of this review is to explore the role of TEs in mammal genome evolution. Below we discuss TE content, advantageous and deleterious effects of TE activity, and the evolution of TE defense strategies, all within a mammalian evolutionary context.

Our knowledge of mammalian genomics is relatively advanced compared to other vertebrates, yet we are still at a point where most of our results are derived from a handful of model taxa. Thus, broad conclusions may reflect clade-specific phenomena rather than generalizations to the entire class of mammals. However, as new sampling methods and sequencing technologies are developed, it will be possible to explore genomes increasing numbers of non-model mammals as well as TE dynamics at the population level to better understand the role of TEs in mammalian evolution.

TE classification

TEs are generally classified into two groups based on their mobilization intermediates (Finnegan 1989). Class I elements, also known as retrotransposons, mobilize as an RNA intermediate. All retrotransposons, commonly called "copy and paste" elements, create new copies of themselves as they are reversely transcribed into the genome. Retrotransposons fall into two major groups, the long terminal repeat (LTR) elements and non-LTR elements, distinguished by the presence or absence of 100–300 bp direct terminal repeats (Fig. 1). The LTR elements, including endogenous retroviruses (ERVs), range in size from a few hundred base pairs to 10 Kb and are structured similarly to retroviruses (discussed below). Autonomous LTR elements encode at least a *gag* and *pol* protein, flanked by the long terminal repeats (LTRs) that give the elements their name. LTR retrotransposons mobilize and replicate through tRNA-primed template switching occurring within a viral-like particle encoded by the *gag* gene (Fig. 1; Leis et al. 1993; Levin 1995).

The non-LTR elements include long interspersed elements (LINEs) and short interspersed elements (SINEs). Both LINEs and SINEs can be identified by the presence of a repetitive tail, usually poly-A, and a lack of LTRs. LINEs are 4–7 Kb long and may encode between one and three proteins that provide the enzymatic machinery necessary for mobilization. The most common mammalian LINE, LINE1 (L1) contains two open reading frames (ORFs), a nuclear chaperone protein (ORF1) and a reverse transcriptase (ORF2; Fig. 1). A third very short protein (ORF0) was recently described in primate L1 elements but its function is unknown (Denli et al. 2015). SINEs can range in size from 150 to 500 bp and lack the machinery necessary for self-mobilization, i.e., non-autonomous (Fig. 1). Most mammal SINEs are derived from the combination of a 5′ head that is derived from a ribosomal or tRNA pseudogene and a 3′ tail homologous to a LINE. The LINE-like region of the SINE is used to parasitize the enzymatic machinery of LINEs for mobilization (Eickbush 1992). Unlike LINEs, de novo origination of SINEs is relatively common in mammals (Fig. 2).

Phylogenetic relationships estimated from the conserved residues of the RT domain indicate four distinct clades including LINE-like retrotransposons, Penelope-like retortransposons, prokaryotic retroelements (ex. group II introns) and the LTR containing retroelements, including LTR retrotransposons and retroviruses (Gladyshev and Arkhipova 2011). Retroviruses are structurally similar to LTR retrotransposons except for the addition of an env gene. The env gene codes an envelope protein which allows a retrovirus to potentially infect other cells. Env gene acquisition could be from
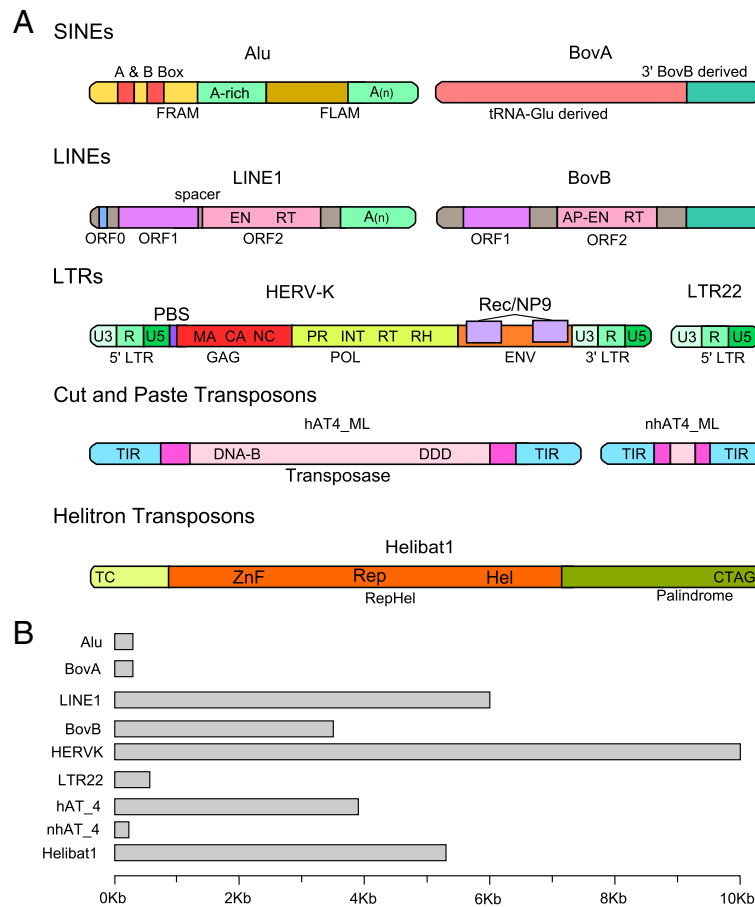
Fig. 1 Mammalian transposable elements. **a** Structure of common mammalian transposable elements. A and B box, promoter regions derived from 7SL RNA; FRAM, free right Alu monomer; FLAM, free left Alu monomer; A(n) poly A repeat; UTR, untranslated region; ORF0, primate-specific open reading frame 0; ORF1, nuclear chaperone protein; ORF2 reverse transcriptase; EN, endonuclease domain; RT, reverse transcriptase domain; AP-EN, apurinic-apyrimidinic endonuclease; U3, unique 3′ sequence; R, repeated sequence; U5, unique 5′ sequence; PBS, tRNA primer binding site; GAG, GAG protein; MA, matrix domain; CA, capsid domain; NC, nucleocapsid domain; POL, polyprotein; PR, protease domain; INT, integrase domain; RH, RNAse H domain; ENV, envelope protein; Rec/NP9, Rec and NP9 proteins including possible alternative splicing events; TIR, terminal inverted repeat; DNA-B, DNA binding domain; DDD, three conserved aspartate residues; TC, TC dinucleotide sequence; ZnF, zinc-finger-containing motifs; RepHel, replicase protein; Rep, replicase domain; Hel, helicase domain; CTAG -CTAG nucleotide sequence. **b** Representative elements drawn to scale

recombination between infected hosts with active retrotransposition of LTR retrotransposons or the acquisition of modification of a host encoded gene (Eickbush and Malik 2002; Koonin et al. 1991). These events have likely occurred multiple times leading to the independent origins of retroviruses with some groups obtaining the ability to leave the cell and infect others including the caulimoviruses and gypsy viruses (Eickbush and Jamburuthugoda 2008; Herédia et al. 2004). Once a germ cell is infected, the retrovirus becomes endogenized and can then be transmitted vertically from parent to offspring.

Class II elements, also known as the DNA transposons, mobilize as a DNA intermediate associated with a transposase. DNA transposons can be subdivided into two major groups; the cut-and-paste and rolling-circle transposons (Wicker et al. 2007). Cut-and-paste elements excise themselves from the genome as a double-stranded DNA intermediate associated with a transposase, an enzyme encoded by autonomous instances of the element family in question. Common cut-and-paste transposons found in mammals include the Tc1/*mariner*, *h*ATs, and *piggyBac* families, all of which can be characterized by their terminal inverted repeats–ranging in

Origination Events
1 Formation of composite SVA SINEs
2 Formation of composite LAVA SINEs
3 Dimerization of FLAM, FRAM to form ALU SINEs
4 Dimeriztaion of ID and B1 to form ID-B1 SINEs
5 De novo origination of FLAM SINEs
6 De novo origination of FRAM SINEs
7 De novo origination of B2 SINEs
8 De novo origination of B3 SINEs
9 De novo origination of Ped-1 SINEs
10 De novo origination of ID SINEs
11 De novo origination of pB1 SINEs
12 De novo origination of ID-Spe SINEs
13 De novo origination of CAN SINEs
14 De novo origination of CHR SINEs
15 De novo origination of VES SINEs
16 De novo origination of MEG SINEs
17 De novo origination of AfroSINE SINEs
18 De novo origination of snoRTE SINEs
19 De novo origination of Mon-1 SINEs
Horizontal Transfer Events
20 Horizontal transfer of SPIN transposons
21 Horizontal transfer of hAT transposons
22 Horizontal transfer of piggyBac transposons
23 Horizontal transfer of Tc1/Mariner transposons
24 Horizontal transfer of Helitrons
25 Horizontal transfer of BovB RTE LINEs
Reduction or Expansion Events
26 Reduced B1 SINE activity
27 Reduced LINE activity
28 Reduced activity of LINE2 LINEs
29 Reduced activity of MIR SINEs
30 Reduced LINE1 and SINE activity
31 Maintenence of two LINE1 lineages
Ancestral Elements
32 CR1 LINEs present in ancestral genome
33 LINE1 LINEs present in ancestral genome
34 LINE2 LINEs present in ancestral genome
35 MIR SINEs present in ancestral genome
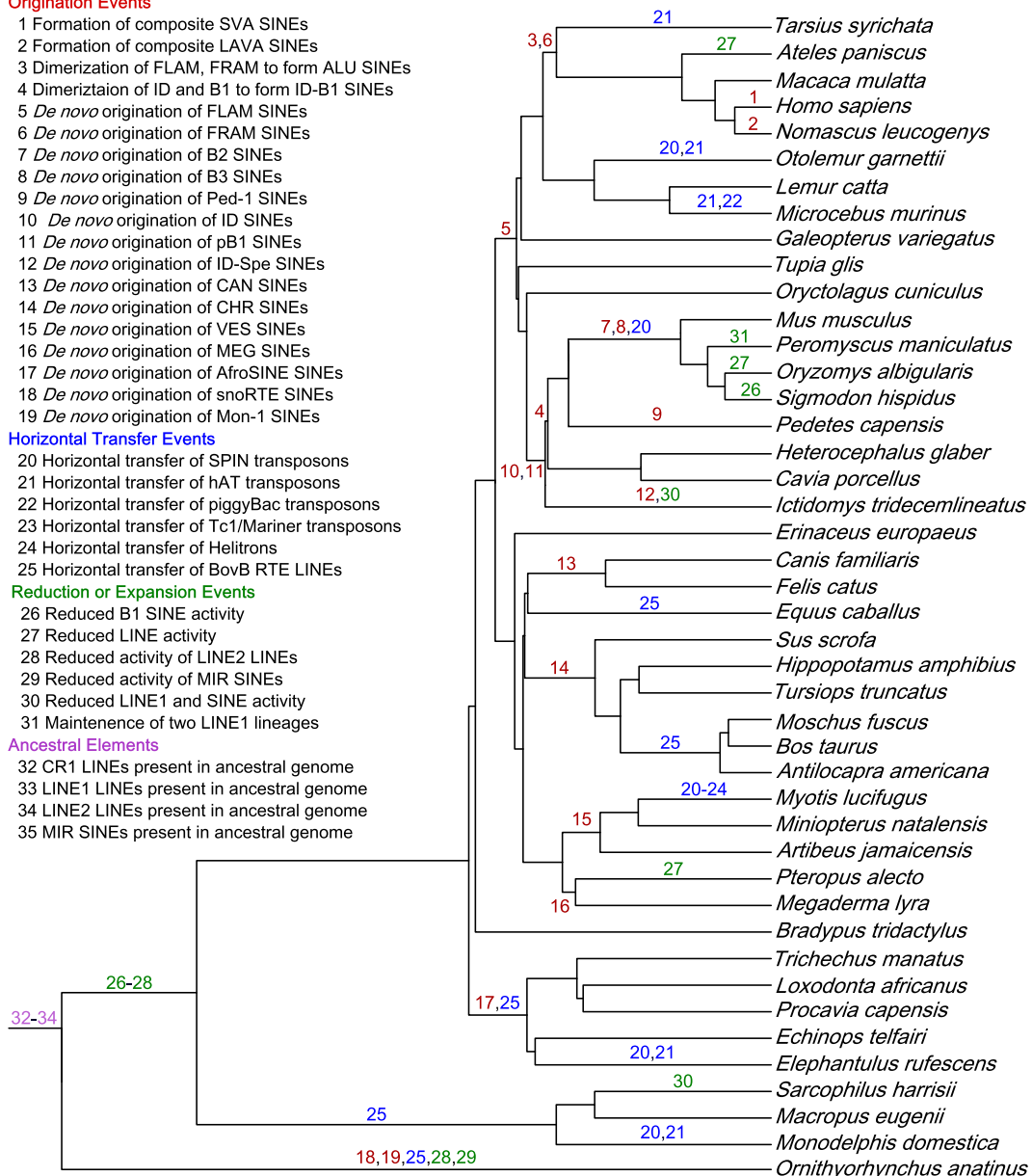
Fig. 2 Major transitions in TE content along are plotted along the mammalian phylogeny. The mammal phylogeny is modified from Meredith et al. (Meredith et al. 2011). Events were inferred to specific nodes using information from (Alföldi et al. 2011; Churakov et al. 2010; Gogolevsky et al. 2008; Gogolevsky et al. 2009; Green et al. 2014; Hillier et al. 2004; Kriegs et al. 2007; Lupan et al. 2015; Nikaido et al. 2003; Novick et al. 2010; Pace et al. 2008; Pagan et al. 2010; Pritham and Feschotte 2007; Ray et al. 2015; Ray et al. 2006; Rinehart et al. 2005; Shimamura et al. 1999; Smit et al. 1995; Suh et al. 2014; Vassetzky and Kramerov 2002; Walsh et al. 2013; Wang et al. 2005; Warren et al. 2008) and are generally classified into four categories. "Origination Events" (red) refer to the de novo or composite origin of new TEs. "Horizontal Transfer Events" (blue) refer to the horizontal transmission of TEs from non-mammalian lineages. "Reduction or Expansion Events" (green) refer to dramatic shifts in accumulation patterns. "Ancestral Elements" (pink) refer to elements that were present in the ancestral mammalian genome

size from 10 to 1,000 bp–and the catalytic domains of their transposase (Fig. 1; Feschotte and Pritham 2007). Rolling-circle transposons, or *Helitrons*, mobilize as a single-stranded DNA copying itself via rolling-circle replication (Kapitonov and Jurka 2001). Autonomous *Helitrons* contain a RepHel protein and a ~20 bp

palindrome that functions as a termination sequence 10–20 bp from the 3′ end of the element (Fig. 1; Kapitonov and Jurka 2007).

Mammalian TE content and evolution

A survey of TEs from species spanning the mammalian phylogeny (Meredith et al. 2011) can be used to generalize TE content in mammals as a whole (Fig. 2). Under this assumption, we can make the following observations regarding TE content and dynamics in mammalian genomes:

- One half to one third of the mammal genome is derived from TEs.
- LINE and SINE retrotransposons are the most common types of TE. DNA transposons are rare and/or ancient.
- Mammals usually have one or more actively mobilizing TE family
- Horizontal transfer of TEs is rare in mammals

As with any observations made across a group as diverse as mammals and a phenomenon as dynamic as TEs, there are exceptions to each of these observations. Data supporting the above observations plus the exceptions are described below.

One half to one third of the genome is derived from TEs

TEs typically make up between one third and one half of mammal genomes (Fig. 3; Elsik et al. 2009; Lander et al. 2001; Mikkelsen et al. 2007; Miller et al. 2008; Warren et al. 2008; Waterston and Pachter 2002) but it is likely that estimates of TE content are biased downwards based on computational and methodological limitations of TE identification. At the time of insertion, novel TE insertions are identical, or nearly identical to, the parent insertion. As neutrally evolving TE insertions age, the genetic distance between insertions increases. Ancient insertions with less than 50% sequence similarity to a query TE rapidly become unidentifiable using homology-based methods. In addition to natural sequence divergence between TE insertions, homology-based searches may not identify lineage-specific repeats if they are not defined a priori (Platt II et al. 2016a). De novo TE identifications tend to

be more accurate and can be used to identify particularly ancient elements. For comparison, current estimates of TE content in the human genome range from 49% to as high as 69% when using homology or de novo-based searches, respectively (de Koning et al. 2011). Because of sequence degradation and homology-based limitations, estimates of TE content are always biased against older elements and sometimes against newer elements. It is almost certain that mammalian genomes are more repeat rich than currently recognized.

LINEs and SINEs are more abundant than LTR elements and DNA transposons

The repetitive portion of mammal genomes is dominated by LINEs and SINEs, followed by LTR retrotransposons, and then DNA transposons. In most, ~ 75% of the repetitive portion is derived from non-LTR retrotransposons (Fig. 3; Lander et al. 2001; Waterston and Pachter 2002). Non-LTR retrotransposons in the platypus genome make up 97.4% of all repetitive sequences and LINEs by themselves occupy 20% of the genome (Warren et al. 2008). The LINE-1 (L1) family is the most successful TE family in mammals, and frequently occupies hundreds of megabases in therian genomes. SINE expansions in mammals piggyback on the success of their autonomous LINE partners. Rather than continuous expansion from a single SINE family, as in the case of L1, unique and lineage-specific SINE families have arisen multiple times (Fig. 2; Kramerov and Vassetzky 2011).

LTR retrotransposons are present in mammal genomes at moderately high copy number, and can occupy between ~4 and 10% of the genome (Mikkelsen et al. 2007). However, the accumulation of LTR elements in the genome may not reflect past activity because recombination between the terminal repeats of LTR elements can remove nearly the entire element, leaving behind a solitary LTR (Bennetzen and Kellogg 1997; Smit 1993).

DNA transposons are usually present in low copy numbers relative to retrotransposons, occupying less than 3% of mammalian genomes (Platt II and Ray 2012). Low copy number of DNA transposons is driven in part by two factors. First, most mammals lack autonomous class II elements (Pace and Feschotte 2007), so the DNA transposons that are present in mammalian genomes are decaying vestiges of earlier transposition events. Second, if a cut-and-paste DNA
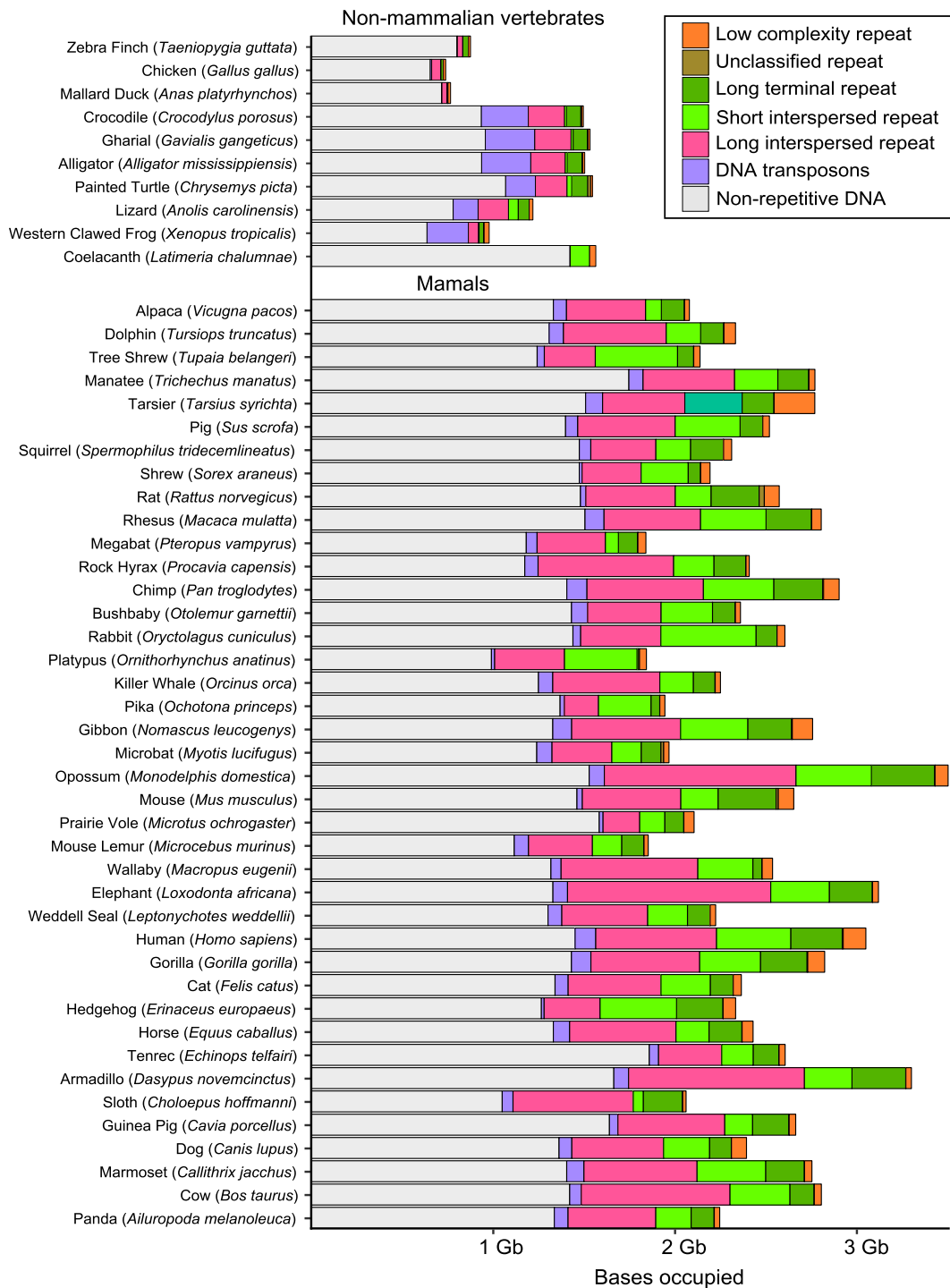
Fig. 3 TE content in mammal and non-mammal vertebrate genomes. TE content was quantified from pre-masked genomes available at http://repeatmasker.org/genomicDatasets/RMGenomicDatasets.html (last accessed 30 November 2014). Genome size was estimated from the number of bases in the genome assembly

transposon is active, a new copy is only created if it reinserts in front of a replication fork. To date, the

vespertilionid bats are the only mammals with significant, active DNA transposition (discussed below).

Most mammals have one or more actively mobilizing TE family

The vast majority of TE insertions in mammal genomes are incapable of mobilization (Fanning 1983). With L1 elements, this is true primarily because most insertions are truncated at the 5′ end due to inefficient reverse transcription (Grimaldi et al. 1984). Furthermore, de novo insertions of any type may be mutated during the insertion process or targeted for transcriptional and post-transcriptional silencing by one of several defense mechanisms (described below). So, while there may be tens of thousands of copies of any TE present, only a small fraction is capable of mobilizing at any moment. In humans and mice, only 6 of more than ~868,000 and 2382 of ~599,000 L1s are retrocompetent in their respective genomes (Brouha et al. 2003; Zemojtel et al. 2007).

Despite the limited number of retrocompetent elements, retrotransposition of LINEs and/or SINEs persists in mouse, human, and most other mammal genomes examined to date. TE quiescence of LINEs and/or SINES has only been observed in the ground squirrel (Platt II and Ray 2012), Tasmanian devil (Nilsson 2016), *Ateles* spider monkeys (Boissinot et al. 2004), sigmodontine rodents (Grahn et al. 2005; Rinehart et al. 2005), and pteropodid bats (Cantrell et al. 2008) but, given the timing and phylogenetic distribution of these silencing events, it is plausible that quiescence has impacted in as many as 15% of all mammal species (Platt II and Ray 2012). Understanding the genomic mechanisms, population genetic parameters, and random factors that reduce or eliminate TE activity is one of the major questions in vertebrate genome evolution (Goodier 2016).

Horizontal transfer of TEs is relatively rare

TEs in mammals are spread primarily through vertical inheritance, though the mobile nature of TEs means the insertion patterns may vary among lineages. In some rare instances, TEs are horizontally transferred among taxa. In mammals, there are fewer than 20 documented horizontal transfer (HT) events in the last 160 MY, compared to 2248 in the last 10 MY in insects (Fig. 2; Peccoud et al. 2017). The most successful HT event in mammals was the expansion of BovB LINES into a diverse group of mammals including afrotherians (ex. elephants and tenerecs), ruminants (ex. cattle and deer), marsupials (ex. kangaroos and possums), and

protherians (platypuses and echidnas). The expansion of BovB in mammals is the result of at least four separate transfer events (Walsh et al. 2013) as recently as 50 MYA (Kordis and Gubensek 1998) likely from a parasite vector (Walsh et al. 2013). Despite the relatively recent invasion into an ancestral ruminant, BovB LINEs and SINEs make up as much as 18.4% (Elsik et al. 2009) and 10.37% (Ge et al. 2013) of the bovid genome, respectively.

Various DNA transposon families have successfully invaded mammalian genomes but not accumulated as successfully as the BovBs. HT of SPIN transposons occurred between 15 and 46 MYA into the galago, murine rodents, opossum, tenerec, and vespertilionid bats (Pace et al. 2008), *hAT*s were transferred to the opossum, tenerec, some primates, and vespertilionid bats (Novick et al. 2010) and *piggyBac*s were transferred into the mouse lemur and vesperitionlind bats (Pagan et al. 2010). The frequency of these events and the fact that many occur in the same taxa suggest that some species may be more susceptible to HT than others, most notably the vespertilionid bats, who also have experienced HT of *Helitrons*, and Tc1/*mariners*. The HT of *hAT*s, *piggyBac*s, Tc1/*mariner*s, and *Helitron* transposons into these bats occurred over a series of events within a narrow 10–20 MY window (Platt II et al. 2016b; Pritham and Feschotte 2007; Ray et al. 2008; Ray et al. 2006; Thomas et al. 2011). As a result, almost half of all recent transposition in the vesper bats has been from DNA transposons compared to less than 1% in other laurasiatherians (Platt II et al. 2014). Still, DNA transposons only account for 3–5% of vespertilionid genomes despite the recent increase in accumulation (Pagán et al. 2012).

Impacts of advantageous TE insertions on mammalian genomes

The presence of TEs and their dynamic nature has shaped mammal genomes in significant ways (reviewed in Chalopin et al. 2015; Sotero-Caio et al. 2017; Warren et al. 2015). Below we describe some advantageous and deleterious effects of TE activity and content. We emphasize that despite our focus on the selectively advantageous or deleterious impacts of TEs, many TE insertions have accumulated through non-adaptive processes associated with reduced effective population sizes and

the resulting increased effects of genetic drift (Lynch and Conery 2003).

Exaptation of TEs

More than a quarter million conserved non-coding elements are derived from TEs in the human genome alone (Lowe and Haussler 2012). Further, identification of additional exaptation events, notably TEs as regulatory units, is becoming increasingly common. Some TE's promoters contain transcription factor binding sites and other regulatory motifs (Bourque et al. 2008). As TEs mobilize in the genome, they spread their own regulatory motifs to new loci. If a TE inserts near-promoter regions of other genes, selection can then co-opt the TE's regulatory elements to alter gene expression of the nearby genes. Given enough time, a novel regulatory network can emerge where a single transcription factor, originally associated with TE transcription, may link dozens, or even hundreds, of previously unrelated genes (Chuong et al. 2017). As selection acts on the nascent network, it can become highly specialized. In the case of MER20, binding sites for hormone responsiveness and pregnancy-related transcription factors found within the TE itself were spread throughout the genome of the ancestral placental mammal.

Multiple examples exist including the differentiation of endometrial cells in the presence of progesterone, which was a critical step in the evolution of pregnancy in placental mammals. Thirteen percent of genes associated with differentiation of endometrial stromal cells appear to be regulated by motifs found in a eutherian-specific *hAT* transposon, MER20. In addition, almost half of all MER20s in the human genome are found within 200 Kb of progesterone responsive genes (Lynch et al. 2011). In some instances, MER20 would insert next genes providing transcription factor-binding sites (or epigenetic modifications) that may not have previously affected that genes. As this continued to happen, a complex network of genes, partially regulated by MER20-derived regulatory sites, developed into a cell-type specific regulatory network for differentiation of endometrial cells (Lynch et al. 2011; Lynch et al. 2015). In another example, three TEs, an AmnSINE, X6b_DNA transposon, and MER117 hAT inserted adjacent to each other in a sequential manner to form a complex promoter for secondary palate development in eutherian mammals. None of the insertions exhibit promoter activity on their own but instead work

cooperatively to regulate *wnt5* expression (Nishihara et al. 2016). Chuong et al. (2017) provides a more detailed review of TE-driven regulatory networks.

Co-option of LTRs is inversely proportional to the age of the element subfamily; younger elements are more likely to be co-opted, a trend that contrasts with the co-option rate of other TEs (Franke et al. 2017). LTRs can serve as gene-remodeling platforms where the promoters and initial exons of a gene are derived from an LTR. LTR-derived promoters and 5′ exons are incorporated into 842 protein coding and lncRNA genes expressed during the transition from oocyte to zygote in rodents (Franke et al. 2017). In mice an MT-C LTR insertion into a DICER intron has truncated the first 6 exons, provided an alternative promoter and novel first exon. This DICER isoform has acquired oocyte-specific expression and is essential for fertility (Flemr et al. 2013).

TEs may promote adaptability

TEs can alter gene expression, disrupt coding genes, transduce exons, or promote recombination allowing for dramatic and rapid restructuring of the genome that may exceed the changes offered by point mutations. These changes may allow populations to more fully explore a fitness landscape in a shorter period of time; increasing the "adaptability" of the population (Casacuberta and González 2013). The role of TEs in promoting adaptability has been explored theoretically (Werren 2011), in the laboratory (Stoebel and Dorman 2010), observed in the wild (Schrader et al. 2014), and has become a critical framework to understand invasion genetics (Stapley et al. 2015), but has yet to be demonstrated in a mammalia. Two hypotheses regarding the role of TEs in promoting adaptability are directly associated with mammals; the stress-response (McClintock 1984) and TE Thrust hypotheses (Oliver and Greene 2011). A variant of the stress-response hypothesis, the epi-transposon hypothesis (Zeh et al. 2009), posits that during times of environmental stress epigenetic suppression of TEs is relaxed allowing for burst of TE activity. The increased rates of TE activity allow populations to explore the fitness landscape. The epi-transposon hypothesis has been explored more completely in plant studies (for examples see Ito et al. 2016; Nozawa et al. 2017) than in vertebrates. Despite these limitations general observations in the human

genome, including increases in TE expression due to chemical exposure, approximating environmental stress (Kale et al. 2005), and the association of TEs in stress related gene regions (van de Lagemaat et al. 2003), tend to support tenets of the epi-transposons hypothesis. The TE-Thrust hypothesis proposes that lineages with TE activity are more fecund than those without, and is based on the observation that TE accumulation tends to be associated with novel genetic change. Like the epi-transposon hypothesis, it is difficult to directly test the TE-Thrust hypothesis. Instead, correlations between TE activity with major evolutionary innovations in mammalian lineages represent the best supporting evidence (Brandt et al. 2005; Lowe and Haussler 2012; Mikkelsen et al. 2007; Pace and Feschotte 2007; Platt II et al. 2014; Suh et al. 2014). However, given that TE activity is more likely to lead to declines in fitness reductions rather than increases in fitness, a rigorous test of the assumption that TE activity leads to as an example, increased diversification rates, to be tested within a phylogenetic context.

Impacts of deleterious TEs insertions on mammalian genomes

Around 10% of all de novo mutations in lab mice are the result of TE insertions (Maksakova et al. 2006). In fact, the mutagenic power of TEs is so great that transposons are often used to identify gene function in model organisms including humans and mice (Dupuy et al. 2005). The deleterious effects of TE activity can result in reduced fitness in populations. Below we discuss the evidence of and known deleterious effects of TEs in mammals, most of which is derived from studies in human and mouse models.

Selection drives TE distribution

One difficulty in identifying deleterious insertions is that they are often lethal or only slightly deleterious and present no obvious phenotypes. As a result, our knowledge of deleterious TE insertions is more limited than one might expect given the frequency with which deleterious TE insertions are expected to arise. Rather than directly observing the deleterious effects at the morphological level, the deleterious nature of

TE insertions can be inferred from biases in their distribution across the genome.

Surveys of the human genome show that TE insertions are not randomly distributed. Younger retrotransposon insertions are biased toward AT rich regions reflecting their target-site preference (Medstrand et al. 2002). As Alu elements age, their density in GC rich regions increases relative to L1 (Jurka et al. 2004). In vertebrates, GC rich regions are typically associated with high gene density compared to the relatively gene poor AT rich regions. Accumulation of elements in GC rich regions likely reflects selection against inter-TE recombination, the only mechanism thought to remove TEs, since such events are more likely to disrupt coding regions than to salvage them (Abrusán and Krambeck 2006; for an alternative explanation see Kvikstad and Makova 2010; Medstrand et al. 2002). If young insertions are neutral or slightly deleterious, selection should remove the slightly deleterious insertions. Under this assumption, older elements are more likely to be neutral than their younger counterparts. Because of GC-biased accumulation, TEs are found in introns of almost 90% of human and mouse genes. Intronic TE insertions tend to be located more than 150 bp away from the closest exons and in the opposite orientation of the gene (Burns and Boeke 2012). These positional biases reflect selection against insertions that disrupt splice sites and/or generate anti-sense exonic transcripts via read-through transcription from promoters in the TE insertion. Interestingly, most known mutagenic insertions found in introns violate these patterns (Zhang et al. 2011).

TE contribution to mutational meltdown

While some levels of TE activity may promote adaptation (discussed above), it is possible that excess TE activity can contribute to mutational meltdown of populations. Mutational meltdown is a positive feedback loop where deleterious mutations accumulate in populations leading to decreased fitness and reduced population sizes which are more prone to the accumulation of additional deleterious mutations via drift. Mutational meltdown is difficult to observe directly because it is a feed-forward loop that results in extinction. As a result, the literature on mutational meltdown is biased toward theoretical work (Lynch et al. 1995) or experiments with lab populations (Zeyl et al. 2001), and less is known about the process in wild species with long generation

times (Rowe and Beebee 2003). The best example of mutational meltdown in mammals is from the Wrangel Island mammoth genome. The single genome available for this group shows an accumulation of gene deletions, premature stop codons, and reduced heterozygosity compared to other mammoth genomes from individuals in larger populations (Rogers and Slatkin 2017). In addition, the Wrangel Island mammoth has an excess of retrogenes. Because retrogenes are a direct byproduct of LINE activity, their presence indicates a burst of retrotransposon activity just prior to extinction of the Wrangle Island mammoth population (Rogers and Slatkin 2017). This final burst of TE activity, beyond creating retrogenes, could have contributed to the mutational meltdown of the Wrangel Island mammoth population. Interestingly, this burst of TE activity is expected under the epi-transposon hypothesis (Zeh et al. 2009).

Somatic diseases

TE expression was previously thought to be limited to the germ line but recent studies identified extensive TE expression in somatic tissue. When occurring in the germ line, lethal insertions are purged, but somatic stem cell insertions are more tolerable and associated with cancer, neuropathy, and the aging process.

TE insertions have been associated with more than 100 diseases (Chénais 2013; Hancks and Kazazian 2016) including several forms of cancer. TE-induced cancer can arise by altering tumor suppressor genes or proto-oncogenes (Morse et al. 1988). In each case, disruption of one allele by a TE usually needs to coincide with a loss-of-function mutation at the other allele before tumorigenesis, so the prevalence of TE-induced tumorigenesis is unknown (Burns and Boeke 2012). Still, examples of TE-driven tumorigenesis exist. For example, an L1 element insertion into the APC tumor suppressor gene initiates colorectal cancer (Scott et al. 2016). Splice variants of Rec and Np9 from the *env* gene in HERV-Ks bind and suppress the promyelocytic leukemia zinc finger protein which is a transcriptional repressor of the oncogene c-myc (Denne et al. 2007). Changes in genome structure including chromosomal translocations, recombination between, and duplications driven by Alu insertions have all been associated with several forms of leukemia (Jeffs et al. 1998; O'Neil et al. 2007; Strout et al. 1998).

Neural tissues contain unexpected levels of TE expression. More than 2200 somatic TE insertions were identified in just three individuals, a large number of which tended to be within or in close proximity to protein-coding genes (Baillie et al. 2011). It has been estimated that 1 in every 300 neuronal genomes contains a novel L1 insertion (Evrony et al. 2016), which means that the typical mammal brain may contain millions, possibly billions of novel TE insertions. These and other observations seem to imply that some level of TE expression is necessary for normal neuronal development, yet progenitor cells are vulnerable to accumulating deleterious mutations (Li et al. 2012; Reilly et al. 2013). Since TEs are going to insert into open, euchromatin sites, they are likely to insert near, or potentially into, transcribed genes associated with neural function (reviewed in Nekrutenko and Li 2001). Environmental stimuli ranging from exposure to light (deHaro et al. 2014), heavy metals (Kale et al. 2005), aromatic hydrocarbons (Stribinskis and Ramos 2006), even physical exercise (Muotri et al. 2009), can potentially increase L1 expression beyond the already elevated levels in neural tissues. As a result, neural tissues show increased accumulation of TE insertions in genes associated with stress, including alcoholism and post-traumatic stress disorder (Reilly et al. 2013). In humans, Alu elements retrotransposed into the mitochondrial, TOMM40 gene 16 times leading to a serious of conformational changes and/or truncated proteins that are less than fully functional (Larsen et al. 2017). Mitochondrial dysfunction in neural cells likely leads to increased oxidative stress and subsequent inflammatory response imitating a feedforward loop that leads to reduced neural function and disease (Swerdlow and Khan 2004).

Mammalian protection against TEs

Because TEs are capable of compromising genome integrity, disrupting gene function, and inducing disease states the genome has evolved several, semi-redundant defensive systems to limit TE activity. These systems range from transcriptional silencing to transcript editing. Below we present three defense mechanisms in mammalian genomes and discuss how they have shaped mammalian genome evolution.

KRAB/KAP1 histone modification

Retrotransposons are transcriptionally silenced in early embryos by histone modification and DNA methylation, although the initiators of this process have, until

recently, been unknown. Krüpell-associated box (KRAB) domain-containing zinc finger proteins (KRAB-ZFPs) have sequence-specific DNA binding ability via C-terminal zinc fingers (Urrutia 2003) that are used to recognize retrotransposons in early embryos (Rowe et al. 2010). After recognition, KRAB-ZFPs recruit KRAB-associated protein 1 (KAP1), which in turn can bind to any one of a series of epigenetic regulators, including, histone methyltransferases (ESET), heterochromatin protein 1 (HP1), nucleosome remodeling and deactylation (NuRD) complex, and DNMT3A and DNMT3B which methylates DNA (Ecco et al. 2017; Feschotte and Gilbert 2012). Knocking out KAP1 in early mouse embryos leads to an upregulation of ERV retrotransposons (Rowe et al. 2010).

Sequence editing with APOBECs

The APOBEC family of cytidine deaminases act by editing reverse (cDNA) transcripts (Friedli and Trono 2015; Mangeat et al. 2003). The enzymes mediate the deamination of cytosine to uracil, causing either direct destruction or debilitating levels of hypermutation in the TE cDNA (Harris et al. 2003). APOBECs arose in early vertebrates, but at least two subfamilies of APOBECs (APOBEC3 and APOBEC1) arose through gene duplications in early mammals, and APOBEC3 is only present in placental mammals (Conticello 2008; Conticello et al. 2004; Rogozin et al. 2007). The family member APOBEC3G was first noticed to edit viral cDNA from HIV lacking the *vif* gene and murine leukemia virus (Bishop et al. 2004). ERVs are structurally like these retroviruses and are also in fact edited by APOBECs. It was first hypothesized that non-LTR retroelements would not be affected by APOBECs because APOBEC3G is restricted to the cytoplasm and non-LTR reverse transcription occurs in the nucleus. However, it was later discovered that several APOBEC3s are expressed in the nucleus and the family inhibits a broad range of both LTR and non-LTR retroelements (Bogerd et al. 2006; Friedli and Trono 2015; Kinomoto et al. 2007; Richardson et al. 2014; Schumann 2007).

PIWI proteins and piRNAs

The typical mammalian genome encodes four PIWI proteins: PIWIL1 (MIWI), PIWIL2 (MILI), PIWIL3, and PIWIL4 (MIWI2). Their partners, piRNAs, are the most abundant small RNA in testis and range from approximately 24 to 32 bases. piRNAs have few distinguishing characteristics except for a uridine bias in the first position of sense-oriented sequences. PIWI proteins and the associated piRNAs are predominately expressed in the germ line and are required for spermatogenesis (Aravin et al. 2006; Carmell et al. 2007; Kuramochi-Miyagawa et al. 2004; Lau et al. 2006).

PIWIs silence TEs through two pathways, direct cleavage of TE transcripts and de novo methylation of TE loci, both of which are dependent on the "ping-pong" cycle where PIWIs use piRNAs as guides to TE transcripts. Briefly, sense primary piRNAs direct PIWIL2 to complementary anti-sense TE-derived transcripts. These transcripts are cleaved to generate anti-sense secondary piRNAs and these secondary piRNA are incorporated into a PIWIL4 or PIWIL2 complex which is guided to a sense TE transcript, and the cycle is repeated creating a feed forward loop that increases the number of piRNA guides and reduces the abundance of TE mRNAs. During testis development, genome-wide methylation marks are erased and reset in primordial germ cells. As a result, TEs are released from epigenetic silencing and their expression increases (Molaro et al. 2014). During this time, PIWIL4 is thought to mark TE loci for downstream de novo methylation by methyltransferases DNMT3L and DNMT3A (Aravin et al. 2008; Molaro et al. 2014); however, this exact mechanism is unknown. Knocking out PIWIL4 and PIWIL2 lead to upregulation of retrotransposon in the male germ line, an arrest of gametogenesis, and complete sterility in male mice, likely due to unrestricted retrotransposon mobilization (Aravin et al. 2007; Carmell et al. 2007).

As a primary defense mechanism against TE proliferation, PIWI processing likely affects TE composition in mammalian genomes (Vandewege et al. 2016). When comparing TE expression and piRNA processing in mammals with different active TE families, a strong positive relationship between piRNA and TE transcript abundance was present; more piRNAs were derived from highly expressed TEs. The relationships between TE transcription and piRNA quantity, however, did not correlate with "efficiency" in targeting. Young SINE expression in the dog and horse genomes is comparable, yet young SINEs accumulate at a much greater rate in the dog than the horse genome. SINE transcripts in the horse are targeted, cleaved, and/or methylated allowing for high SINE expression but limited accumulation when compared to dogs. This reduced efficiency of PIWI processing in the dog genome has resulted in 166,148 SINEC_Cfa SINE

insertions versus 53,092 ERE1 in the horse over a relatively similar amount of time and despite similar SINE transcription levels (Vandewege et al. 2016). SINE accumulation is so rapid in the dog that ~ 10,000 bimorphic loci in the domestic dog population (Wang and Kirkness 2005).

The genomic arms race with TEs

TEs and genomic defenses systems are engaged in an arms race that mirrors the relationship between pathogens and the immune system. The genome must constantly develop strategies to fight transposition, which pressures TEs to evolve to escape repression. For example, by resembling regulatory sequences, some LTRs likely escape methylation in embryonic stem cells (Gerdes et al. 2016). With the exception of birds, there are hundreds of KRAB-ZFPs encoded by most vertebrate genomes (Emerson and Thomas 2009; Liu et al. 2014). KRAB-ZFPs experience rapid evolutionary changes in zinc finger structure, sequence, and expression, and splicing patterns (Nowick et al. 2010) and tandem duplication events drive the rapid expansion of the KRAB-ZFP gene families. Further, there is a positive correlation between the number and age KRAB-ZFPs and genomes ERV content (Thomas and Schneider 2011). Selection is also a strong driver of evolution in APOBEC sequences (Sawyer et al. 2004). APOBEC3G has been under strong positive selection in primates, and additional members of the APOBEC family display strong signals of positive selection in humans (Sawyer et al. 2004). Primate genomes encode the most APOBECs, and the expansion of this family during primate evolution coincides with a decrease in TE activity (Schumann 2007). The PIWIs are fundamentally different from APOBECs and KRAB-ZFPs, where selection is a strong driver of their evolution. The sequences and structure of PIWI proteins are well conserved, but the targeting mechanism, the piRNAs are directly processed from active TEs allowing the inhering targeting and silencing of the newest and most expressed TEs (Molaro et al. 2014; Vandewege et al. 2016).

The genomic TE defense system is overall adaptive and redundant given there are defenses at every stage of the TE replication cycle. TEs are silenced via KRAB-ZFPs during embryogenesis and PIWI proteins methylate and silence young TEs during testis development. Elements that escape methylation and become expressed are cleaved in the cytoplasm by additional PIWI proteins. And if a TE mRNA is not cleaved, APOBECs edit the transcript so that the new insertion

is no longer functional. These pathways, and others just being described (Martinez et al. 2017; Schorn et al. 2017), work together to prevent full-length autonomous elements from propagating in mammalian genomes.

Conclusion

One third to one half of the typical mammalian genome is derived from TEs, primarily non-LTR retrotransposons. Because of their abundance, TEs can have significant impacts on mammalian genome evolution. Active (retro)transposition can provide opportunities for expatiation events, build novel regulatory networks, and even increase the adaptive potential of a population. Despite these benefits, many insertions with any phenotypic impact are neutral or deleterious. Highly deleterious insertions will be rapidly purged from the gene pool; however, somatic insertions can arise presenting as any number of cancers or neurological diseases. To mitigate the potential deleterious effects of TE activity, several redundant defense mechanisms have evolved to limit TE expression.

Our knowledge of mammalian genomics is relatively advanced compared to other vertebrates, yet we are still at a point where most of our results are derived from a handful of model taxa. At this point, broad conclusions may reflect clade-specific phenomena rather than generalize to the entire class of mammals. As new sampling methods and sequencing technologies are developed, it will become possible to explore genomes of non-model, mammalian tax at the population level to truly understand the role of TEs in mammalian evolution.

References

Abrusán G, Krambeck HJ (2006) The distribution of L1 and Alu retroelements in relation to GC content on human sex chromosomes is consistent with the ectopic recombination model. J Mol Evol 63(4):484–492. https://doi.org/10.1007/s00239-005-0275-0

Alföldi J, di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, Ray DA, Boissinot S, Shedlock AM, Botka C, Castoe TA, Colbourne JK, Fujita MK, Moreno RG, ten Hallers BF, Haussler D, Heger A, Heiman D, Janes DE, Johnson J, de Jong PJ, Koriabine MY, Lara M, Novick PA, Organ CL, Peach SE, Poe S, Pollock DD, de Queiroz K, Sanger T, Searle S, Smith JD, Smith Z, Swofford R, Turner-Maier J, Wade J, Young S, Zadissa A, Edwards SV, Glenn TC, Schneider CJ, Losos JB, Lander ES, Breen M, Ponting CP, Lindblad-Toh K (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477(7366):587–591. https://doi.org/10.1038/nature10390

Aravin A et al (2006) A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442:203

Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318(5851):761–764. https://doi.org/10.1126/science.1146484

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 31(6):785–799. https://doi.org/10.1016/j.molcel.2008.09.003

Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, de Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, Talbot RT, Gustincich S, Freeman TC, Mattick JS, Hume DA, Heutink P, Carninci P, Jeddeloh JA, Faulkner GJ (2011) Somatic retrotransposition alters the genetic landscape of the human brain. Nature 479(7374):534–537. https://doi.org/10.1038/nature10531

Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genomic obesity? Plant Cell 9(9):1509–1514. https://doi.org/10.1105/tpc.9.9.1509

Biémont C (2010) A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186(4):1085–1093. https://doi.org/10.1534/genetics.110.124180

Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho S-J, Malim MH (2004) Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol 14(15):1392–1396. https://doi.org/10.1016/j.cub.2004.06.057

Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR (2006) APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res 34(1):89–95. https://doi.org/10.1093/nar/gkj416

Boissinot S, Roos C, Furano AV (2004) Different rates of LINE-1 (L1) retrotransposon amplification and evolution in New World monkeys. J Mol Evol 58(1):122–130. https://doi.org/10.1007/s00239-003-2539-x

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18(11):1752–1762. https://doi.org/10.1101/gr.080663.108

Brandt J, Schrauth S, Veith AM, Froschauer A, Haneke T, Schultheis C, Gessler M, Leimeister C, Volff JN (2005) Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. Gene 345(1):101–111. https://doi.org/10.1016/j.gene.2004.11.022

Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH (2003) Hot L1s account for the bulk of retrotransposition in the human population. P Natl Acad Sci USA 100(9):5280–5285. https://doi.org/10.1073/pnas.0831042100

Burns KH, Boeke JD (2012) Human transposon tectonics. Cell 149(4):740–752. https://doi.org/10.1016/j.cell.2012.04.019

Cantrell MA, Scott L, Brown CJ, Martinez AR, Wichman HA (2008) Loss of LINE-1 activity in the megabats. Genetics 178(1):393–404. https://doi.org/10.1534/genetics.107.080275

Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell 12(4):503–514. https://doi.org/10.1016/j.devcel.2007.03.001

Casacuberta E, González J (2013) The impact of transposable elements in environmental adaptation. Mol Ecol 22(6):1503–1517. https://doi.org/10.1111/mec.12170

Chalopin D, Naville M, Plard F, Galiana D, Volff J-N (2015) Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol 7(2):567–580. https://doi.org/10.1093/gbe/evv005

Chen J, Greenblatt IM, Dellaporta SL (1992) Molecular analysis of Ac transposition and DNA replication. Genetics 130(3):665–676

Chénais B (2013) Transposable elements and human cancer: a causal relationship? Biochim Biophys Acta, Rev Cancer 1835(1):28–35. https://doi.org/10.1016/j.bbcan.2012.09.001

Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Pääbo S, Rocchi M, Eichler EE (2005) A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437(7055):88–93. https://doi.org/10.1038/nature04000

Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18:71

Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J (2010) Rodent evolution: back to the root. Mol Biol Evol 27(6):1315–1326. https://doi.org/10.1093/molbev/msq019

Conticello SG (2008) The AID/APOBEC family of nucleic acid mutators. Genome Biol 9(6):229. https://doi.org/10.1186/gb-2008-9-6-229

Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS (2004) Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. Mol Biol Evol 22:367–377

Cowley M, Oakey RJ (2013) Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet 9(1):e1003234. https://doi.org/10.1371/journal.pgen.1003234

de Koning AJ, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7(12):e1002384. https://doi.org/10.1371/journal.pgen.1002384

deHaro D, Kines KJ, Sokolowski M, Dauchy RT, Streva VA, Hill SM, Hanifin JP, Brainard GC, Blask DE, Belancio VP (2014) Regulation of L1 expression and retrotransposition by melatonin and its receptor: implications for cancer risk associated with light exposure at night. Nucleic Acids Res 42(12):7694–7707. https://doi.org/10.1093/nar/gku503

Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MCN, Diedrich JK, Aslanian A, Ma J, Moresco JJ, Moore L, Hunter T, Saghatelian A, Gage FH (2015) Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell 163(3):583–593. https://doi.org/10.1016/j.cell.2015.09.025

Denne M, Sauter M, Armbruester V, Licht JD, Roemer K, Mueller-Lantzsch N (2007) Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. J Virol 81(11):5607–5616. https://doi.org/10.1128/JVI.02771-06

Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature 436(7048): 221–226. https://doi.org/10.1038/nature03691

Ecco G, Imbeault M, Trono D (2017) KRAB zinc finger proteins. Development 144(15):2719–2729. https://doi.org/10.1242/dev.132605

Eickbush T (1992) Transposing without ends: the non-LTR retrotransposable elements. New Biol 4(5):430–440

Eickbush TH, Jamburuthugoda VK (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134(1-2):221–234. https://doi.org/10.1016/j.virusres.2007.12.010

Eickbush TH, Malik HS (2002) Origins and evolution of retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington D.C., pp. 1111–1144.

Elsik CG, Tellam RL, Worley KC (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324(5926):522–528. https://doi.org/10.1126/science.1169588

Emerson RO, Thomas JH (2009) Adaptive evolution in zinc finger transcription factors. PLoS Genet 5(1):e1000325. https://doi.org/10.1371/journal.pgen.1000325

Evrony GD, Lee E, Park PJ, Walsh CA (2016) Resolving rates of mutation in the brain using single-neuron genomics. Elife 5: e12966

Fanning TG (1983) Size and structure of the highly repetitive BAM HI element in mice. Nucleic Acids Res 11(15):5073–5091. https://doi.org/10.1093/nar/11.15.5073

Feschotte C, Gilbert C (2012) Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet 13(4):283–296. https://doi.org/10.1038/nrg3199

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41(1):331–368. https://doi.org/10.1146/annurev.genet.40.110405.090448

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5(4):103–107. https://doi.org/10.1016/0168-9525(89)90039-5

Flemr M, Malik R, Franke V, Nejepinska J, Sedlacek R, Vlahovicek K, Svoboda P (2013) A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell 155(4):807–816. https://doi.org/10.1016/j.cell.2013.10.001

Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J, Svobodova E, Ma J, Suzuki Y, Aoki F, Schultz RM, Vlahovicek K, Svoboda P (2017) Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res 27:1384–1394. https://doi.org/10.1101/gr.216150.116

Friedli M, Trono D (2015) The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Bi 31(1):429–451. https://doi.org/10.1146/annurev-cellbio-100814-125514

Ge R-L, Cai Q, Shen YY, San A, Ma L, Zhang Y, Yi X, Chen Y, Yang L, Huang Y, He R, Hui Y, Hao M, Li Y, Wang B, Ou X, Xu J, Zhang Y, Wu K, Geng C, Zhou W, Zhou T, Irwin DM, Yang Y, Ying L, Bao H, Kim J, Larkin DM, Ma J, Lewin HA, Xing J, Platt RN, Ray DA, Auvil L, Capitanu B, Zhang X, Zhang G, Murphy RW, Wang J, Zhang YP, Wang J (2013) Draft genome sequence of the Tibetan antelope. Nat Commun 4:1858. https://doi.org/10.1038/ncomms2860

Gerdes P, Richardson SR, Mager DL, Faulkner GJ (2016) Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. Genome Biol 17(1): 100. https://doi.org/10.1186/s13059-016-0965-5

Gladyshev EA, Arkhipova IR (2011) A widespread class of reverse transcriptase-related cellular genes. P Natl Acad Sci USA 108(51):20311–20316. https://doi.org/10.1073/pnas.1100266108

Gogolevsky KP, Vassetzky NS, Kramerov DA (2008) Bov-B-mobilized SINEs in vertebrate genomes. Gene 407(1-2):75–85. https://doi.org/10.1016/j.gene.2007.09.021

Gogolevsky KP, Vassetzky NS, Kramerov DA (2009) 5S rRNA-derived and tRNA-derived SINEs in fruit bats. Genomics 93(5):494–500. https://doi.org/10.1016/j.ygeno.2009.02.001

Goodier JL (2016) Restricting retrotransposons: a review. Mob DNA 7(1):16. https://doi.org/10.1186/s13100-016-0070-z

Grahn R, Rinehart T, Cantrell M, Wichman H (2005) Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. Cytogenet Genome Res 110(1-4):407–415. https://doi.org/10.1159/000084973

Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, Vandewege MW, St. John JA, Capella-Gutierrez S, Castoe TA, Kern C, Fujita MK, Opazo JC, Jurka J, Kojima KK, Caballero J, Hubley RM, Smit AF, Platt RN, Lavoie CA, Ramakodi MP, Finger JW, Suh A, Isberg SR, Miles L, Chong AY, Jaratlerdsiri W, Gongora J, Moran C, Iriarte A, McCormack J, Burgess SC, Edwards SV, Lyons E, Williams C, Breen M, Howard JT, Gresham CR, Peterson DG, Schmitz J, Pollock DD, Haussler D, Triplett EW, Zhang G, Irie N, Jarvis ED, Brochu CA, Schmidt CJ, McCarthy FM, Faircloth BC, Hoffmann FG, Glenn TC, Gabaldon T, Paten B, Ray DA (2014) Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346(6215):1254449. https://doi.org/10.1126/science.1254449

Grimaldi G, Skowronski J, Singer MF (1984) Defining the beginning and end of KpnI family segments. EMBO J 3:1753

Hancks DC, Kazazian HH (2016) Roles for retrotransposon insertions in human disease. Mob DNA 7(1):9. https://doi.org/10.1186/s13100-016-0065-9

Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH (2003) DNA deamination mediates innate immunity to retroviral infection. Cell 113(6):803–809. https://doi.org/10.1016/S0092-8674(03)00423-9

Herédia F, Loreto ELS, Valente VLS (2004) Complex evolution of gypsy in Drosophilid species. Mol Biol Evol 21(10):1831–1842. https://doi.org/10.1093/molbev/msh183

Hillier LW et al (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate

evolution. Nature 432(7018):695–716. https://doi.org/10.1038/nature03154

Ito H et al (2016) A stress-activated transposon in Arabidopsis induces transgenerational abscisic acid insensitivity. Sci Rep 6:srep23181

Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM (1998) The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. Hum Mol Genet 7(5):767–776. https://doi.org/10.1093/hmg/7.5.767

Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV (2004) Duplication, coclustering, and selection of human Alu retrotransposons. P Natl Acad Sci USA 101(5):1268–1272. https://doi.org/10.1073/pnas.0308084100

Kale SP, Moore L, Deininger PL, Roy-Engel AM (2005) Heavy metals stimulate human LINE-1 retrotransposition. Int J Env Res Pub He 2(1):14–23. https://doi.org/10.3390/ijerph2005010014

Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. P Natl Acad Sci USA 98(15):8714–8719. https://doi.org/10.1073/pnas.151269298

Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet 23(10):521–529. https://doi.org/10.1016/j.tig.2007.08.004

Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K (2007) All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res 35(9):2955–2964. https://doi.org/10.1093/nar/gkm181

Koonin EV, Mushegian AR, Ryabov EV, Dolja VV (1991) Diverse groups of plant RNA and DNA viruses share related movement proteins that may possess chaperone-like activity. J Gen Virol 72(12):2895–2903. https://doi.org/10.1099/0022-1317-72-12-2895

Kordis D, Gubensek F (1998) Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. P Natl Acad Sci USA 95(18):10704–10709. https://doi.org/10.1073/pnas.95.18.10704

Kramerov D, Vassetzky N (2011) Origin and evolution of SINEs in eukaryotic genomes. Heredity 107(6):487–495. https://doi.org/10.1038/hdy.2011.43

Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J (2007) Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. Trends Genet 23(4):158–161. https://doi.org/10.1016/j.tig.2007.02.002

Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N, Fujita Y, Ikawa M, Iwai N, Okabe M, Deng W, Lin H, Matsuda Y, Nakano T (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development 131(4):839–849. https://doi.org/10.1242/dev.00973

Kvikstad EM, Makova KD (2010) The (r) evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. Genome Res 20(5):600–613. https://doi.org/10.1101/gr.099044.109

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowski J (2001) Initial sequencing of the human genome. Nature 409:860–921. https://doi.org/10.1038/35057062

Larsen PA, Lutz MW, Hunnicutt KE, Mihovilovic M, Saunders AM, Yoder AD, Roses AD (2017) The Alu neurodegeneration hypothesis: a primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease. Alzheimers Dement 13(7):828–838. https://doi.org/10.1016/j.jalz.2017.01.017

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE (2006) Characterization of the piRNA complex from rat testes. Science 313(5785):363–367. https://doi.org/10.1126/science.1130164

Leis J, Aiyar A, Cobrinik D (1993) Regulation of initiation of reverse transcription of retroviruses. Cold Spring Harbor M 23:33–33

Levin HL (1995) A novel mechanism of self-primed reverse transcription defines a new family of retroelements. Mol

Cell Biol 15(6):3310–3317. https://doi.org/10.1128/MCB.15.6.3310

Li W, Jin Y, Prazak L, Hammell M, Dubnau J (2012) Transposable elements in TDP-43-mediated neurodegenerative disorders. PLoS One 7(9):e44099. https://doi.org/10.1371/journal.pone.0044099

Liu H, Chang L-H, Sun Y, Lu X, Stubbs L (2014) Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. Genome Biol Evol 6(3):510–525. https://doi.org/10.1093/gbe/evu030

Lowe CB, Haussler D (2012) 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. PLoS One 7(8):e43128. https://doi.org/10.1371/journal.pone.0043128

Lupan I, Bulzu P, Popescu O, Damert A (2015) Lineage specific evolution of the VNTR composite retrotransposon central domain and its role in retrotransposition of gibbon LAVA elements. BMC Genomics 16(1):389. https://doi.org/10.1186/s12864-015-1543-z

Lynch M, Conery JS (2003) The origins of genome complexity. Science 302(5649):1401–1404. https://doi.org/10.1126/science.1089370

Lynch M, Conery J, Burger R (1995) Mutation accumulation and the extinction of small populations. Amer Nat 146(4):489–518. https://doi.org/10.1086/285812

Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet 43(11):1154–1159. https://doi.org/10.1038/ng.917

Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, Emera D, Sheikh SZ, Grützner F, Bauersachs S, Graf A, Young SL, Lieb JD, DeMayo FJ, Feschotte C, Wagner GP (2015) Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep 10(4):551–561. https://doi.org/10.1016/j.celrep.2014.12.052

Maksakova IA, Romanish MT, Gagnier L, Dunn CA, Van de Lagemaat LN, Mager DL (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet 2(1):e2. https://doi.org/10.1371/journal.pgen.0020002

Mangeat B, Turelli P, Caron G, Friedli M (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424(6944):99–103. https://doi.org/10.1038/nature01709

Marques-Bonet T, Girirajan S, Eichler EE (2009) The origins and impact of primate segmental duplications. Trends Genet 25(10):443–454. https://doi.org/10.1016/j.tig.2009.08.002

Martinez G, Choudury SG, Slotkin RK (2017) tRNA-derived small RNAs target transposable element transcripts. Nucleic Acids Res 45(9):5142–5152. https://doi.org/10.1093/nar/gkx103

McClintock B (1984) The significance of responses of the genome to challenge. Science 226(4676):792–801. https://doi.org/10.1126/science.15739260

Medstrand P, Van De Lagemaat LN, Mager DL (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res 12(10):1483–1495. https://doi.org/10.1101/gr.388902

Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simao TLL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL, Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, Murphy WJ (2011) Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334(6055):521–524. https://doi.org/10.1126/science.1211028

Mi S, Lee X, Xiang-ping L, Veldman GM (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403(6771):785–789. https://doi.org/10.1038/35001608

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Baldwin J, Abdouelleil A, Abdulkadir J, Abebe A, Abera B, Abreu J, Acer SC, Aftuck L, Alexander A, An P, Anderson E, Anderson S, Arachi H, Azer M, Bachantsang P, Barry A, Bayul T, Berlin A, Bessette D, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Bourzgui I, Brown A, Cahill P, Channer S, Cheshatsang Y, Chuda L, Citroen M, Collymore A, Cooke P, Costello M, D'Aco K, Daza R, de Haan G, DeGray S, DeMaso C, Dhargay N, Dooley K, Dooley E, Doricent M, Dorje P, Dorjee K, Dupes A, Elong R, Falk J, Farina A, Faro S, Ferguson D, Fisher S, Foley CD, Franke A, Friedrich D, Gadbois L, Gearin G, Gearin CR, Giannoukos G, Goode T, Graham J, Grandbois E, Grewal S, Gyaltsen K, Hafez N, Hagos B, Hall J, Henson C, Hollinger A, Honan T, Huard MD, Hughes L, Hurhula B, Husby ME, Kamat A, Kanga B, Kashin S, Khazanovich D, Kisner P, Lance K, Lara M, Lee W, Lennon N, Letendre F, LeVine R, Lipovsky A, Liu X, Liu J, Liu S, Lokyitsang T, Lokyitsang Y, Lubonja R, Lui A, MacDonald P, Magnisalis V, Maru K, Matthews C, McCusker W, McDonough S, Mehta T, Meldrim J, Meneus L, Mihai O, Mihalev A, Mihova T, Mittelman R, Mlenga V, Montmayeur A, Mulrain L, Navidi A, Naylor J, Negash T, Nguyen T, Nguyen N, Nicol R, Norbu C, Norbu N, Novod N, O'Neill B, Osman S, Markiewicz E, Oyono OL, Patti C, Phunkhang P, Pierre F, Priest M, Raghuraman S, Rege F, Reyes R, Rise C, Rogov P, Ross K, Ryan E, Settipalli S, Shea T, Sherpa N, Shi L, Shih D, Sparrow T, Spaulding J, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Strader C, Tesfaye S, Thomson T, Thoulutsang Y, Thoulutsang D, Topham K, Topping I, Tsamla T, Vassiliev H, Vo A, Wangchuk T, Wangdi T, Weiand M, Wilkinson J, Wilson A, Yadav S, Young G, Yu Q, Zembek L, Zhong D, Zimmer A, Zwirko Z, Jaffe DB, Alvarez P, Brockman W, Butler J, Chin CW, Gnerre S, MacCallum I, Graves JAM, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh K (2007) Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature 447(7141):167–177. https://doi.org/10.1038/nature05805

Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A, Tikhonov A, Raney B, Patterson N, Lindblad-Toh K, Lander ES, Knight JR, Irzyk

GP, Fredrikson KM, Harkins TT, Sheridan S, Pringle T, Schuster SC (2008) Sequencing the nuclear genome of the extinct woolly mammoth. Nature 456(7220):387–390. https://doi.org/10.1038/nature07446

Molaro A, Falciatori I, Hodges E, Aravin AA, Marran K, Rafii S, McCombie WR, Smith AD, Hannon GJ (2014) Two waves of de novo methylation during mouse germ cell development. Genes Dev 28(14):1544–1549. https://doi.org/10.1101/gad.244350.114

Morse B, Rotherg PG, South VJ, Spandorfer JM, Astrin SM (1988) Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. Nature 333(6168):87–90. https://doi.org/10.1038/333087a0

Muotri AR, Zhao C, Marchetto MC, Gage FH (2009) Environmental influence on L1 retrotransposons in the adult hippocampus. Hippocampus 19(10):1002–1007. https://doi.org/10.1002/hipo.20564

Nekrutenko A, Li W-H (2001) Transposable elements are found in a large number of human protein-coding genes. Trends Genet 17(11):619–621. https://doi.org/10.1016/S0168-9525(01)02445-3

Nikaido M, Nishihara H, Hukumoto Y, Okada N (2003) Ancient SINEs from African endemic mammals. Mol Biol Evol 20(4):522–527. https://doi.org/10.1093/molbev/msg052

Nilsson MA (2016) The devil is in the details: transposable element analysis of the Tasmanian devil genome. Mobile Genet Elem 6(1):e1119926. https://doi.org/10.1080/2159256X.2015.1119926

Nishihara H, Kobayashi N, Kimura-Yoshida C, Yan K, Bormuth O, Ding Q, Nakanishi A, Sasaki T, Hirakawa M, Sumiyama K, Furuta Y, Tarabykin V, Matsuo I, Okada N (2016) Coordinately co-opted multiple transposable elements constitute an enhancer for wnt5a expression in the mammalian secondary palate. PLoS Genet 12(10):e1006380. https://doi.org/10.1371/journal.pgen.1006380

Novick P, Smith J, Ray D, Boissinot S (2010) Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. Gene 449(1-2):85–94. https://doi.org/10.1016/j.gene.2009.08.017

Nowick K, Hamilton AT, Zhang H, Stubbs L (2010) Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. Mol Biol Evol 27(11):2606–2617. https://doi.org/10.1093/molbev/msq157

Nozawa K, Kawagishi Y, Kawabe A, Sato M, Masuta Y, Kato A, Ito H (2017) Epigenetic regulation of a heat-activated retrotransposon in cruciferous vegetables. Epigenomes 1(1):7. https://doi.org/10.3390/epigenomes1010007

Ohno S (1970) Evolution by gene duplication. Springer, New York. https://doi.org/10.1007/978-3-642-86659-3

Oliver KR, Greene WK (2011) Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. Mob DNA 2(1):8. https://doi.org/10.1186/1759-8753-2-8

O'Neil J, Tchinda J, Gutierrez A, Moreau L, Maser RS, Wong KK, Li W, McKenna K, Liu XS, Feng B, Neuberg D, Silverman L, DeAngelo DJ, Kutok JL, Rothstein R, DePinho RA, Chin L, Lee C, Look AT (2007) Alu elements mediate MYB gene tandem duplication in human T-ALL. J Exp Med 204(13):3059–3066. https://doi.org/10.1084/jem.20071637

Ostertag EM, Kazazian HH Jr (2001) Biology of mammalian L1 retrotransposons. Annu Rev Genet 35(1):501–538. https://doi.org/10.1146/annurev.genet.35.102401.091032

Pace JK, Feschotte C (2007) The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 17(4):422–432. https://doi.org/10.1101/gr.5826307

Pace JK, Gilbert C, Clark MS, Feschotte C (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. P Natl Acad Sci USA 105(44):17023–17028. https://doi.org/10.1073/pnas.0806548105

Pagan HJ, Smith JD, Hubley RM, Ray DA (2010) PiggyBac-ing on a primate genome: novel elements, recent activity and horizontal transfer. Genome Biol Evol 2(0):293–303. https://doi.org/10.1093/gbe/evq021

Pagán HJ, Macas J, Novák P, McCulloch ES, Stevens RD, Ray DA (2012) Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. Genome Biol Evol 4(4):575–585. https://doi.org/10.1093/gbe/evs038

Peccoud J, Loiseau V, Cordaux R, Gilbert C (2017) Massive horizontal transfer of transposable elements in insects. P Natl Acad Sci USA 114(18):4721–4726. https://doi.org/10.1073/pnas.1621178114

Platt RN II, Ray DA (2012) A non-LTR retroelement extinction in *Spermophilus tridecemlineatus*. Gene 500(1):47–53. https://doi.org/10.1016/j.gene.2012.03.051

Platt RN II, Vandewege MW, Kern C, Schmidt CJ, Hoffmann FG, Ray DA (2014) Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. Mol Biol Evol 31(6):1536–1545. https://doi.org/10.1093/molbev/msu112

Platt RN II, Blanco-Berdugo L, Ray DA (2016a) Accurate transposable element annotation is vital when analyzing new genome assemblies. Genome Biol Evol 8(2):403–410. https://doi.org/10.1093/gbe/evw009

Platt RN II, Mangum SF, Ray DA (2016b) Pinpointing the vesper bat transposon revolution using the *Miniopterus natalensis* genome. Mob DNA 7(1):12. https://doi.org/10.1186/s13100-016-0071-y

Pritham EJ, Feschotte C (2007) Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. P Natl Acad Sci USA 104(6):1895–1900. https://doi.org/10.1073/pnas.0609601104

Ray DA, Pagan HJ, Thompson ML, Stevens RD (2006) Bats with hAT s: evidence for recent DNA transposon activity in genus *Myotis*. Mol Biol Evol 24(3):632–639. https://doi.org/10.1093/molbev/msl192

Ray DA, Feschotte C, Pagan HJ, Smith JD, Pritham EJ, Arensburger P, Atkinson PW, Craig NL (2008) Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. Genome Res 18:717–728. https://doi.org/10.1101/gr.071886.107

Ray DA, Pagan HJ, Platt RN II, Kroll AR, Schaack S, Stevens RD (2015) Differential SINE evolution in vesper and non-vesper bats. Mob DNA 6(1):10. https://doi.org/10.1186/s13100-015-0038-4

Reilly MT, Faulkner GJ, Dubnau J, Ponomarev I, Gage FH (2013) The role of transposable elements in health and diseases of the central nervous system. J Neurosci 33(45):17577–17586. https://doi.org/10.1523/JNEUROSCI.3369-13.2013

Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV (2014) APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. Elife 3:e02008

Rinehart T, Grahn R, Wichman H (2005) SINE extinction preceded LINE extinction in sigmodontine rodents: implications for retrotranspositional dynamics and mechanisms. Cytogenet Genome Res 110(1-4):416–425. https://doi.org/10.1159/000084974

Rogers RL, Slatkin M (2017) Excess of genomic defects in a woolly mammoth on Wrangel island. PLoS Genet 13(3):e1006601. https://doi.org/10.1371/journal.pgen.1006601

Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, Aravind L, Pancer Z (2007) Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat Immunol 8(6):647–656. https://doi.org/10.1038/ni1463

Rowe G, Beebee TJ (2003) Population on the verge of a mutational meltdown? Fitness costs of genetic load for an amphibian in the wild. Evolution 57(1):177–181. https://doi.org/10.1111/j.0014-3820.2003.tb00228.x

Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB, Trono D (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463(7278):237–240. https://doi.org/10.1038/nature08674

Sawyer SL, Emerman M, Malik HS (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol 2(9):e275. https://doi.org/10.1371/journal.pbio.0020275

Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R (2017) LTR-retrotransposon control by tRNA-derived small RNAs. Cell 170(1):61–71.e11. https://doi.org/10.1016/j.cell.2017.06.013

Schrader L, Kim JW, Ence D, Zimin A, Klein A, Wyschetzki K, Weichselgartner T, Kemena C, Stökl J, Schultner E, Wurm Y, Smith CD, Yandell M, Heinze J, Gadau J, Oettler J (2014) Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun 5:5495. https://doi.org/10.1038/ncomms6495

Schumann G (2007) APOBEC3 proteins: major players in intracellular defence against LINE-1 mediated retrotransposition. Biochem Soc Trans. 35:637–642. https://doi.org/10.1042/BST0350637

Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, Devine SE (2016) A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res 26(6):745–755. https://doi.org/10.1101/gr.201814.115

Shimamura M, Abe H, Nikaido M, Ohshima K, Okada N (1999) Genealogy of families of SINEs in cetaceans and artiodactyls: the presence of a huge superfamily of tRNA (Glu)-derived families of SINEs. Mol Biol Evol 16(8):1046–1060. https://doi.org/10.1093/oxfordjournals.molbev.a026194

Smit AF (1993) Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res 21(8):1863–1872. https://doi.org/10.1093/nar/21.8.1863

Smit AF, Tóth G, Riggs AD, Jurka J (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J Mol Biol 246(3):401–417. https://doi.org/10.1006/jmbi.1994.0095

Sotero-Caio CG, Platt RN II, Suh A, Ray DA (2017) Evolution and diversity of transposable elements in vertebrate genomes. Genome Biol Evol 9(1):161–177. https://doi.org/10.1093/gbe/evw264

Stapley J, Santure AW, Dennis SR (2015) Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol 24(9):2241–2252. https://doi.org/10.1111/mec.13089

Stoebel DM, Dorman CJ (2010) The effect of mobile element IS 10 on experimental regulatory evolution in Escherichia coli. Mol Biol Evol 27(9):2105–2112. https://doi.org/10.1093/molbev/msq101

Stribinskis V, Ramos KS (2006) Activation of human long interspersed nuclear element 1 retrotransposition by benzo (a) pyrene, an ubiquitous environmental carcinogen. Cancer Res 66(5):2616–2620. https://doi.org/10.1158/0008-5472.CAN-05-3478

Strout MP, Marcucci G, Bloomfield CD, Caligiuri MA (1998) The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia. P Natl Acad Sci USA 95(5):2390–2395. https://doi.org/10.1073/pnas.95.5.2390

Suh A, Churakov G, Ramakodi MP, Platt RN 2nd, Jurka J, Kojima KK, Caballero J, Smit AF, Vliet KA, Hoffmann FG, Brosius J, Green RE, Braun EL, Ray DA, Schmitz J (2014) Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. Genome Biol Evol 7(1):205–217. https://doi.org/10.1093/gbe/evu256

Swerdlow RH, Khan SM (2004) A "mitochondrial cascade hypothesis" for sporadic Alzheimer's disease. Med Hypotheses 63(1):8–20. https://doi.org/10.1016/j.mehy.2003.12.045

Thomas JH, Schneider S (2011) Coevolution of retroelements and tandem zinc finger genes. Genome Res 21(11):1800–1812. https://doi.org/10.1101/gr.121749.111

Thomas J, Sorourian M, Ray D, Baker RJ, Pritham EJ (2011) The limited distribution of Helitrons to vesper bats supports horizontal transfer. Gene 474(1-2):52–58. https://doi.org/10.1016/j.gene.2010.12.007

Urrutia R (2003) KRAB-containing zinc-finger repressor proteins. Genome Biol 4(10):231. https://doi.org/10.1186/gb-2003-4-10-231

van de Lagemaat LN, Landry J-R, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 19(10):530–536. https://doi.org/10.1016/j.tig.2003.08.004

Vandewege MW, Platt RN II, Ray DA, Hoffmann FG (2016) Transposable element targeting by piRNAs in Laurasiatherians with distinct transposable element histories. Genome Biol Evol 8(5):1327–1337. https://doi.org/10.1093/gbe/evw078

Vassetzky NS, Kramerov DA (2002) CAN—a pan-carnivore SINE family. Mamm Genome 13(1):50–57. https://doi.org/10.1007/s00335-001-2111-1

Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL (2013) Widespread horizontal transfer of retrotransposons. P Natl Acad Sci USA 110(3):1012–1016. https://doi.org/10.1073/pnas.1205856110

Wang W, Kirkness EF (2005) Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res 15(12):1798–1808. https://doi.org/10.1101/gr.3765505

...

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA (2005) SVA elements: a hominid-specific retroposon family. J Mol Biol 354(4):994–1007. https://doi.org/10.1016/j.jmb.2005.09.085

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, López-Otín C, Ordóñez GR, Eichler EE, Chen L, Cheng Z, Deakin JE, Alsop A, Thompson K, Kirby P, Papenfuss AT, Wakefield MJ, Olender T, Lancet D, Huttley GA, Smit AF, Pask A, Temple-Smith P, Batzer MA, Walker JA, Konkel MK, Harris RS, Whittington CM, Wong ES, Gemmell NJ, Buschiazzo E, Vargas Jentzsch IM, Merkel A, Schmitz J, Zemann A, Churakov G, Kriegs JO, Brosius J, Murchison EP, Sachidanandam R, Smith C, Hannon GJ, Tsend-Ayush E, McMillan D, Attenborough R, Rens W, Ferguson-Smith M, Lefèvre CM, Sharp JA, Nicholas KR, Ray DA, Kube M, Reinhardt R, Pringle TH, Taylor J, Jones RC, Nixon B, Dacheux JL, Niwa H, Sekita Y, Huang X, Stark A, Kheradpour P, Kellis M, Flicek P, Chen Y, Webber C, Hardison R, Nelson J, Hallsworth-Pepin K, Delehaunty K, Markovic C, Minx P, Feng Y, Kremitzki C, Mitreva M, Glasscock J, Wylie T, Wohldmann P, Thiru P, Nhan MN, Pohl CS, Smith SM, Hou S, Nefedov M, de Jong PJ, Renfree MB, Mardis ER, Wilson RK (2008) Genome analysis of the platypus reveals unique signatures of evolution. Nature 453(7192):175–183. https://doi.org/10.1038/nature06936

Warren IA, Naville M, Chalopin D, Levin P, Berger CS, Galiana D, Volff J-N (2015) Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosom Res 23(3):505–531. https://doi.org/10.1007/s10577-015-9493-5

Waterston RH, Pachter L (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520–562. https://doi.org/10.1038/nature01262

Werren JH (2011) Selfish genetic elements, genetic conflict, and evolutionary innovation. P Natl Acad Sci USA 108(Supplement_2):10863–10870. https://doi.org/10.1073/pnas.1102343108

Wicker T, Sabot F, Hua-van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973–982. https://doi.org/10.1038/nrg2165

Zeh DW, Zeh JA, Ishida Y (2009) Transposable elements and an epigenetic basis for punctuated equilibria. BioEssays 31(7):715–726. https://doi.org/10.1002/bies.200900026

Zemojtel T, Penzkofer T, Schultz J, Dandekar T, Badge R, Vingron M (2007) Exonization of active mouse L1s: a driver of transcriptome evolution? BMC Genomics 8(1):392. https://doi.org/10.1186/1471-2164-8-392

Zeyl C, Mizesko M, Visser JAGM (2001) Mutational meltdown in laboratory yeast populations. Evolution 55(5):909–917.

Zhang Y, Romanish MT, Mager DL (2011) Distributions of transposable elements reveal hazardous zones in mammalian introns. PLoS Comput Biol 7(5):e1002046. https://doi.org/10.1371/journal.pcbi.1002046