

# Evolution and Diversity of Transposable Elements in Vertebrate Genomes

Cibele G. Sotero-Caio<sup>1,†</sup>, Roy N. Platt II<sup>1,†</sup>, Alexander Suh<sup>2</sup>, and David A. Ray<sup>1,\*</sup>

<sup>1</sup>Department of Biological Sciences, Texas Tech University, Lubbock, TX

<sup>2</sup>Department of Evolutionary Biology (EBC), Uppsala University, Uppsala, Sweden

†These authors contributed equally to this work.

\*Corresponding author: E-mail: david.4.ray@gmail.com.

Accepted: December 6, 2016

## Abstract

Transposable elements (TEs) are selfish genetic elements that mobilize in genomes via transposition or retrotransposition and often make up large fractions of vertebrate genomes. Here, we review the current understanding of vertebrate TE diversity and evolution in the context of recent advances in genome sequencing and assembly techniques. TEs make up 4–60% of assembled vertebrate genomes, and deeply branching lineages such as ray-finned fishes and amphibians generally exhibit a higher TE diversity than the more recent radiations of birds and mammals. Furthermore, the list of taxa with exceptional TE landscapes is growing. We emphasize that the current bottleneck in genome analyses lies in the proper annotation of TEs and provide examples where superficial analyses led to misleading conclusions about genome evolution. Finally, recent advances in long-read sequencing will soon permit access to TE-rich genomic regions that previously resisted assembly including the gigantic, TE-rich genomes of salamanders and lungfishes.

**Key words:** vertebrate, transposable element, transposons, retrotransposons.

## Introduction

In the last decade, high-throughput sequencing has led to substantial advances in the fields of comparative genomics (Berthelot et al. 2014; Louis et al. 2015), systematics (Noonan and McCallion 2010), molecular evolution (McGaugh et al. 2015), functional genomics (Foote et al. 2015), adaptive evolution (Ge et al. 2013), and the evolution of genome architecture (Farré et al. 2015). Although genomic data are still scarce for the majority of vertebrates and restricted mostly to a growing list of model species, increased rates of whole-genome sequencing (from ~26 sequenced vertebrates in 2009 to the 268 that are currently listed at NCBI, <http://www.ncbi.nlm.nih.gov>; last accessed March 1, 2016) have provided insights into overall differences of genome size and composition for representatives of all vertebrate classes (Koepfli et al. 2015). Repetitive sequences, and transposable elements (TEs) in particular, are a major component of vertebrate genomes and contribute to the diversity in genome sizes and structure. Here, we review recent insights into TEs gained from

modern, genome-scale data sets, and how TEs have influenced vertebrate genomes.

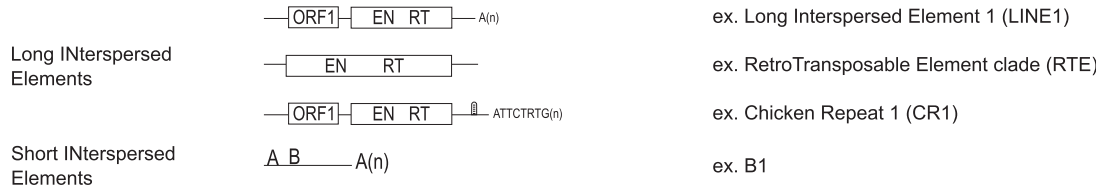
TEs are discrete DNA fragments that have the ability to mobilize within a host genome, often creating new copies of themselves during the mobilization process. Our understanding of these elements has changed dramatically over time. Originally, TEs were thought to be functionless genomic parasites but the complex role they play in genome evolution has received more attention with the increasing availability of genomic data (Orgel and Crick 1980; Oliver and Greene 2009). TEs can impact genome architecture and evolution in numerous ways. They can mediate small-scale changes in linkage groups but also lead to large structural genomic variation, such as deletions, inversions, duplications, and translocations (Gray 2000; Grabundzija, et al. 2016). TEs cause double-strand breaks, which may lead to chromosomal rearrangements, either by TE–TE ectopic recombination or during the transposition process itself (Lim and Simmons 1994; Gray 2000; Hedges and Deininger 2007; Carbone et al. 2014). TEs are also major determinants of genome size and

**Class I - Retrotransposons**

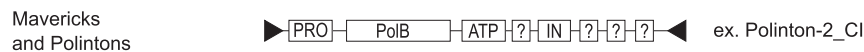
Long Terminal Repeat Retrotransposons



Non-Long Terminal Repeat Retrotransposons



**Class II - DNA Transposons**



**Fig. 1.**—Examples of the major TE types and a general classification scheme. Structural features are presented as follows: LTR retrotransposons – nucleocapsid protein (GAG), envelope protein (ENV), polyprotein (POL) that includes a protease (PRO), integrase (IN), reverse transcriptase (RT), and an RNase H (RH). Non-LTR retrotransposons – nuclear chaperone protein (ORF1), endonuclease domain (EN), RNA polymerase III promoter (A and B), poly-A or other repetitive tail (A(n) or ATTCTRTG(n)). Class II elements – transposase (TPASE), zinc finger domain (ZnF), replicase (REPL), helicase (HELI), polymerase B (PolB), ATPase (ATP), proteins of unknown function (?) and integrase (INT). Coding regions are depicted as boxes. Non-coding regions are depicted as lines. In some instances, the ENV gene in endogenous retroviruses may be missing (dotted line). Triangles represent repeated DNA sequences and the orientation of the triangle reflects the orientation of terminal repeats, if present.

composition in eukaryotes. Indeed, there is a linear correlation between genome size and TE content in all eukaryotes, including vertebrates (Kidwell 2002; Sun, Shepard, et al. 2012; Elliott and Gregory 2015).

TEs are also functionally important. Several putative regulatory sequences derived from TEs are conserved across different vertebrate lineages (Lowe et al. 2007) and are responsible for evolutionary innovation (Levis et al. 1993; Van de Lagemaat et al. 2003; Lowe et al. 2007; Lindblad-Toh et al. 2011; Jurka et al. 2012; Kokošar and Kordiš 2013). Examples include the contribution of TEs to the evolution of the immune system (Kapitonov and Jurka 2005; Chuong, et al. 2016; Lynch 2016), recruitment of a pogo-like transposase into centromeric protein B (CENP-B) (Casola et al. 2008), and transcriptional regulation via non-coding elements in the placenta and brain in mammals (Sasaki et al. 2008; Lynch et al. 2011; Lynch et al. 2015). Altogether, over 200,000 TE insertions have been exapted in the mammalian lineage (Lindblad-Toh, et al. 2011).

**TE Classification, Evolution, and Mobilization**

TEs are split into two major classes (I and II), determined by whether their mode of transposition occurs with or without an RNA intermediate (fig. 1). Each class comprises different

subclasses, superfamilies and families (Finnegan 1989; Wicker et al. 2007; Kapitonov and Jurka 2008; Piégu et al. 2015). Within this broader classification scheme, TEs can be described as those having the ability to self-mobilize (autonomous) and those relying on co-mobilization by the enzymatic machinery of other TEs (non-autonomous).

Class I TEs, or retrotransposons, mobilize in the genomes via a “copy-and-paste” mechanism directed by reverse transcription of an RNA intermediate of a source element. This class is typically subdivided into long terminal repeat (LTR) and non-LTR retrotransposons (Eickbush and Jamburuthugoda 2008). Phylogenetic, structural, and functional differences in the LTR reverse transcriptase seem to indicate a relatively close relationship between LTR retrotransposons, retroviruses and other reverse-transcribing viruses (hepadnaviruses, caulimoviruses), and a more distant kinship with other non-LTRs retrotransposons (Eickbush and Malik 2002).

Although they both employ reverse transcriptase, the two groups of autonomous class I elements, LTR retrotransposons and non-LTR retrotransposons, the latter usually referred to as LINEs (Long Interspersed Elements) mobilize via distinct mechanisms. During LTR retrotransposon mobilization, an RNA molecule is reverse-transcribed into double-stranded DNA through a series of template switches (Levin and Moran 2011). The double-stranded DNA copy of the element is

then reinserted into the genome via an integrase. On the contrary, LINEs mobilize via a target-primed reverse transcription mechanism in which the transcript itself reenters the nucleus with the help of its protein products. Those proteins, usually including an endonuclease domain along with the reverse transcriptase, nick the target site for integration and conduct reverse transcription of the LINE transcript (Luan et al. 1993; Deininger and Batzer 2002).

Non-autonomous non-LTR retrotransposons, also known as Short Interspersed Elements (SINEs), are mobilized by LINE partners, which often share sequence similarity at their 3' ends. The result is that SINEs and LINEs are often found in pairs, where a LINE *trans*-mobilizes a SINE with similar 3' structure (Kajikawa and Okada 2002; Ohshima and Okada 2005). Because the LINE enzymatic machinery is responsible for reverse transcription of the SINE, all that is necessary for a SINE to propagate itself is reliable transcription. For this reason, most SINEs are derived from actively transcribed small RNA genes, such as 5S ribosomal RNA (rRNA), 7SL RNA, or transfer RNA (tRNA), all of which contain internal RNA Pol III promoters. SINEs may acquire 3' sequence similarity to LINEs through the template switching of a reverse transcriptase from a LINE to a tRNA or rRNA during reverse transcription or the insertion of a 5'-truncated LINE into a tRNA or rRNA gene (Ohshima and Okada 2005). Novel SINEs have originated multiple times during vertebrate evolution and include some unusual elements. See, for example, the relatively recent discoveries of the primate-specific SVA SINEs (Wang et al. 2005), or the SINEUs in crocodylians that are derived from U1 and U2 small nuclear RNAs (Kojima 2015).

Class II TEs or DNA transposons mobilize without reverse transcription of source elements and are classified into three major subclasses, each with a distinct transposition mechanism: "cut-and-paste" or Terminal Inverted Repeat (TIR) DNA transposons (e.g., hATs, piggyBacs and mariners), rolling-circle transposons (e.g., Helitrons), and self-synthesizing DNA transposons (e.g., Mavericks) (Kapitonov and Jurka 2001, 2006; Pritham et al. 2007; Bao et al. 2009). Some of these subclasses likely originated from bacterial insertion sequences (IS) (Siguier et al. 2015) or bacteriophages (Krupovic and Koonin 2015). As with the retrotransposons, these elements can also be either autonomous or non-autonomous. In this case, however, the non-autonomous families are often deletion derivatives of their autonomous counterparts (Hartl et al. 1992; Feschotte and Pritham 2007).

Class I elements are found in most eukaryotic lineages and, rarely, in prokaryotes, whereas class II elements are readily found in both prokaryotes and eukaryotes. This suggests that progenitors of both classes were likely present in the common ancestor of all eukaryotes. Despite the ubiquity of TEs in vertebrate genomes, differential amplification of elements in isolated populations, genetic drift, and recombination can result in drastically different genomic TE landscapes between closely related taxa (Bergman and Bensasson 2007;

Akagi et al. 2008; Ray et al. 2008; Jurka et al. 2011; Jurka et al. 2012; Schmidt et al. 2012). Indeed, the differential patterns of TE accumulation and amplification presented by vertebrates have likely played an important role in organismal evolution. TEs can alter or disrupt the expression of genes, promoting population-level variation (Akagi et al. 2008) and possibly even rapid adaptation (Stapley et al. 2015). The evolutionary potential of TE-derived variation has been postulated in the epitransposon hypothesis (Zeh et al. 2009) and the TE-Thrust hypothesis (Oliver and Greene 2011, 2012), and may be strongly influenced by environmental and ecological factors. Indeed, speciation events often correlate with the expansion of new TE families (Boer et al. 2007; Michalak 2009; Oliver and Greene 2009; Zeh et al. 2009; Jurka et al. 2011; Oliver and Greene 2011), suggesting that TEs can serve as drivers of adaptation, diversification, and speciation by generating structural genomic diversity between populations (reviewed in Chénais et al. 2012; Rebollo et al. 2012).

Some TEs are capable of being horizontally transferred (HT) between individuals from distantly related species (Daniels et al. 1990; Gilbert et al. 2010; Pagan et al. 2010; Schaack et al. 2010; Thomas et al. 2010; Gilbert et al. 2013). The ability to horizontally transfer successfully seems to be related to stability of the mobilizing intermediate (Eickbush and Malik 2002). DNA transposons are mobilized as a transposome, a dimeric transposase associated with the double-stranded DNA transposon that is highly stable (Silva et al. 2004). LTR retrotransposons are transcribed into single-stranded RNA but then reverse-transcribed into a double-stranded DNA prior to reintegration, providing a moderately stable intermediate. Non-LTR retrotransposons are mobilized as single-stranded RNA and are relatively unstable (Eickbush and Malik 2002). Following the pattern of stability DNA transposons undergo frequent horizontal transfer, LTR retrotransposons are sometimes horizontally transferred (Schaack et al. 2010), and non-LTR retrotransposons are rarely horizontally transferred (with the notable exception of the retrotransposon-like non-LTR superfamily, RTE; Kordis and Gubensek 1998; Walsh et al. 2013; Suh et al. 2016).

## Annotating TEs

Despite (and possibly because of) the increased availability of genome-scale data, proper and thorough TE annotation is sorely lacking in many of the vertebrate genome projects published in recent years. This observation was recently highlighted in two manuscripts (Hoen et al. 2015; Platt et al. 2016). While many papers focus on the coding regions of genomes or on specific questions related to particular genes, the TE repertoire is often ignored or given only a passing glance. Much of this is due to the fact that a thorough annotation of the TE content in the genome requires significant manual analysis that is not amenable to automation. Such a lack of manual curation on current TE annotation

practices results in inaccurate pictures of the TE landscape that could negatively impact our understanding of those genomes. This was illustrated in Platt et al. (2016) using a variety of mammalian genomes. In that work, the authors demonstrated that by using only homology-based approaches, the TE repertoire is usually underestimated. Furthermore, the divergences of the TEs that are identified tend to be overestimated, leading researchers to believe that TE accumulations are older than they really are.

Identifying repetitive sequences in general, and TEs in particular is computationally challenging, and as mutations accumulate within individual insertions, the difficulty in identification increases. Several tools have been developed to identify TEs through homology, as well as using *de novo* methods (reviewed in Lerat 2010). However, rather than resolving the problem of TE annotation, the large number of tools introduces variability across different annotations as different research groups use their preferred tools. The quality of TE annotations often varies among studies, especially as methodologies continue to improve. For example, 45–69% of the human genome (Lander et al. 2001; de Koning et al. 2011, Wheeler and Eddy 2013) and 25–50% of the coelacanth genome (Amemiya et al. 2013; Nikaido et al. 2013) is derived from TEs, depending on the methodology used (supplementary table S1, Supplementary Material online). These examples emphasize the difficulty not only in TE identification but also in comparing annotations within a species, much less across a group as diverse as vertebrates and demonstrate the need for a standardized annotation methodology or benchmarking metrics (Hoen et al. 2015).

Given the fact that a large number of cellular processes and genomic structural variations are heavily influenced by both the presence and activity of TEs, the inaccurate representation has the potential to influence decision making and hypothesis generation. For example, in their 2012 update to the TE-Thrust hypothesis, Oliver and Greene (2012) reference the genome of the naked mole rat (*Heterocephalus glaber*) as a potential example (Kim et al. 2011). They state that the TEs of the genome “are homogeneous and constitute 25% of the genome, [and] are highly divergent, indicating that they have been both nonviable and inactive for a very long time.” The subsequent analysis by Platt et al. (2016), implementing a more thorough analysis of the TE content, not only identified an additional 4% of the genome as being TE-derived but also found, to paraphrase Mark Twain, that the report of the demise of TEs was greatly exaggerated. Indeed, a relatively recent surge in LINE accumulation was identified. Oliver and Greene further suggested that there may be a link between the lack of recent TE activity and *H. glaber*'s reduced incidence of cancer. However, this hypothesis is a direct result of the lack of accurate information about the TE complement and, for all we know, has led to laboratories pursuing a line of questioning that has no actual support. Such examples highlight the need to keep up with the growing number of genome

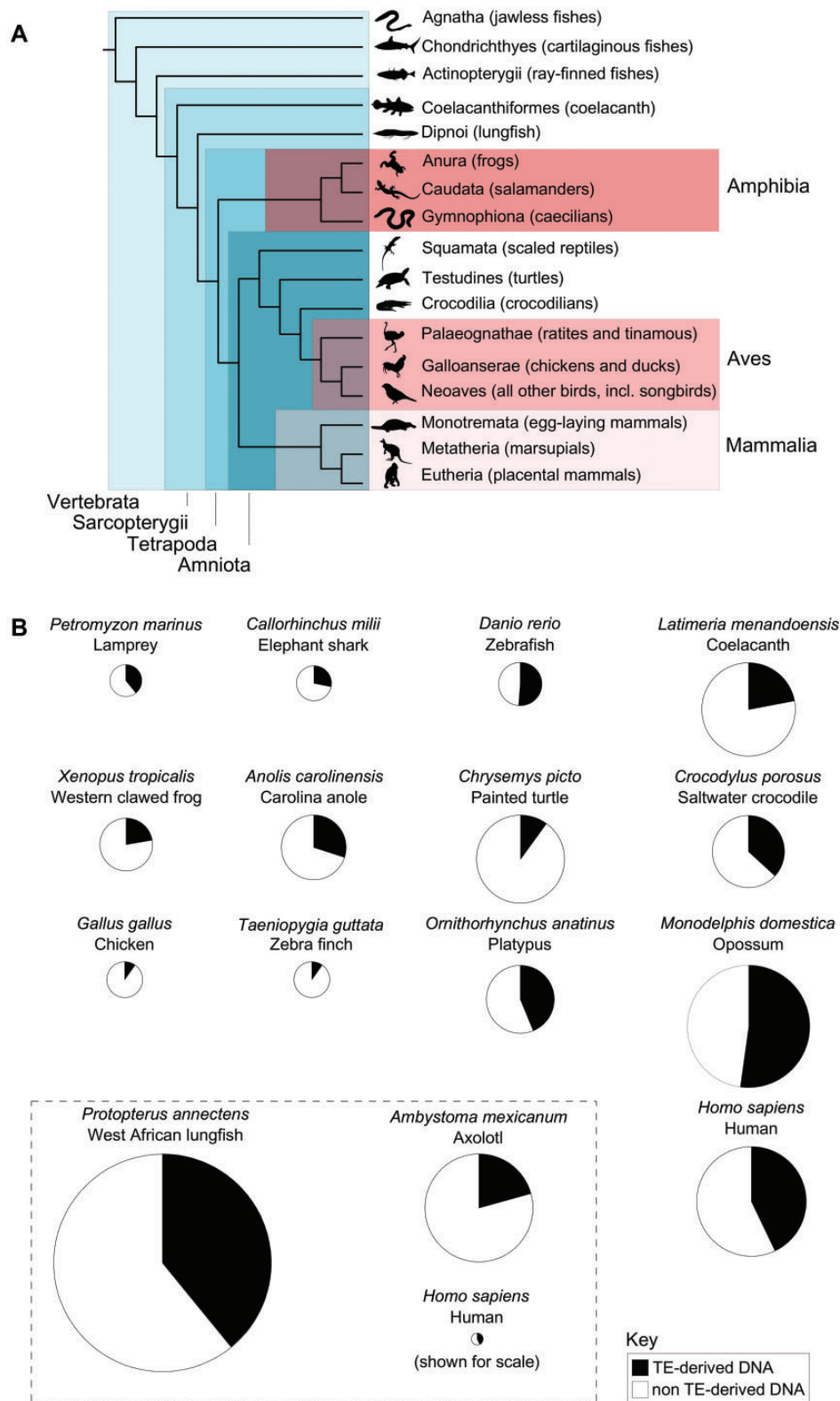
assemblies available to the research community by providing accurate assessments of TE content from both qualitative and quantitative perspectives.

We should also point out that our ability to evaluate the TE content of a genome is directly related to the quality and effort given to TE classification and annotations for that genome. For example, much more effort has been devoted to the accurate classification of TEs in model organisms like human, mouse and *Drosophila*. Indeed, the taxonomy of Alu SINEs is very well described with 68 distinct subfamilies currently described in RepBase and several others present in the literature (Ray and Batzer 2005). By contrast, most SINE families in other vertebrates are rarely even divided into subfamilies. The result can be a lack of precision when identifying and dating TE accumulations in the less well-studied species.

Furthermore, as described by Pop (2009), the primary difficulty in creating an accurate genome assembly is the presence of the TEs themselves. The presence of these repetitive regions leads to breaks in the assemblies and the regions spanning those repeats are likely missing. This is particularly problematic for genomes with recent accumulations. Insertions that are highly similar or very long are more difficult to assemble since similar reads may be collapsed and few read pairs span long repeats.

These problems are particularly well exemplified by recent advances in genome assembly and resulting TE annotations. The study of Alu, L1 and SVA activity in primates, especially humans and fellow hominids, exemplify the problem. Work by various authors has suggested that humans are particularly rich in recent Alu activity compared with chimpanzees, gorillas, and orangutans while SVA elements have experienced an increase in accumulation in the branch leading to the human-chimpanzee common ancestor (Hedges et al. 2004; Mikkelsen et al. 2005; Mills et al. 2006; Eid et al. 2009; Ventura et al. 2011; Lee et al. 2015). Indeed, orangutans appear to have experienced a nearly complete cessation of Alu retrotransposition over the past 12 million years (Locke et al. 2011). Most recently, Gordon et al. (2016) provided an excellent illustration by performing *de novo* assembly of a gorilla genome via long, single-molecule, real-time (SMRT) reads. They demonstrated that substantial portions of TE-derived regions of the original Sanger- and Illumina-based genome assembly were missing. The long-read assembly increased the number of Alu repeats by 3.8-fold and increased the number of identifiable full-length PTERV1 elements, which can encompass 10 kb, by nearly 5-fold. Thus, as is always the case in scientific endeavors, our understanding of TE diversity in genomes will undoubtedly change as our methods of genome assembly improve.

Understanding TE diversity is more important than ever given their presence in virtually all vertebrate genomes, evolutionary impacts, and the increasing volume of genomic data being generated. Below, we review our current understanding of TE biology in each of the major vertebrate lineages (fig. 2A).



**Fig. 2.**—(A) A general phylogeny of vertebrates. (B) Pie charts comparing the percentage of a genome derived from TEs. The area of the pie chart is proportional to genome size except for the west African lungfish and Axolotl. The genomes of these species are exceptionally large. A scaled pie-chart representing the human genome is presented for to illustrate this aspect. Please note that in all cases, the TE content is based on estimates made using the methods by the individual research teams studying each genome. Thus, the methods employed varied and represent the authors’ best estimates obtained using those methods.

We find a general pattern that the larger and more deeply branching the clade, the more variety we see with regard to the diversity of landscapes and TE content (fig. 2B).

## TE Patterns in the Major Vertebrate Lineages

### Fishes

Fishes are herein defined as the paraphyletic group comprising jawless, cartilaginous, ray-finned, and lobe-finned fishes (including coelacanth and lungfish), thus comprising the five deepest lineages of vertebrates (Amemiya et al. 2013). There is a dramatic variation in TE copy number and composition in different fish taxa, ranging from 55% in the zebrafish (*Danio rerio*) to only 6% in the green spotted pufferfish (*Tetraodon nigroviridis*), one of the smallest known vertebrate genomes (Crollius et al. 2000; Volff et al. 2003; Howe et al. 2013). All major types of eukaryotic TEs are present and fishes display an overall higher TE diversity than other vertebrate groups (Chalopin et al. 2014). Most TE annotation efforts to date have targeted ray-finned fishes, but characterization of the TE landscapes have been accomplished for at least one species within each major fish lineage, as detailed below.

### Agnatha

TE information derived from whole-genome sequencing data is available for a single jawless fish species: the sea lamprey (*Petromyzon marinus*). Curation of the lamprey genome assembly suggests that at least 34.7% is derived from TEs (Smith et al. 2013). The genome includes lineage-specific TEs, as well as ancient repeats shared with other vertebrates and even invertebrates. Many of the TEs have yet to be classified and these make up ~19.2% of the genome. Most of the known TEs are estimated to be derived from LINEs and SINEs but class II elements are also present, and to a lesser extent, LTR retrotransposons (Smith et al. 2013; Chalopin et al. 2015). The lamprey genome also contains thousands of copies Tc1 transposons with high sequence similarity (reaching 92–98% identity) to those found in diverse lineages of teleost fishes, suggesting both recent amplification of these elements and high rates of HT (Kuraku et al. 2012). A similar scenario was also described for Chapaev transposons (Zhang et al. 2014). Host–parasite interactions between teleosts and lampreys may, therefore, play a role in mediating horizontal transfers of TEs. Interestingly, lampreys eliminate ~20% of their somatic genome during embryogenesis, much of it TE-rich (Smith et al. 2009, 2013). It is unknown how this feature might contribute to a distinctive TE dynamic. For example, one could argue that the elimination of junk in the form of TEs in somatic cells is a strategy for tolerating TEs in the population as a whole, while also preserving TE-derived variation in the germ line. Reducing genome complexity and streamlining cell reproduction or other biological processes in the “everyday life” of somatic

cells may increase host fitness. In any case, it is well possible that programmed DNA elimination is present in all jawless fishes and might be even more widespread among vertebrates (Wang and Davis 2014).

### Chondrichthyes

Only a single-sequenced genome is available for cartilaginous fishes (Chondrichthyes) (enkaresh, et al. 2014). The elephant shark (*Callorhynchus milii*) harbors a TE content of ~42% of its ~770-Mb to 1-Gb genome. The prevalent TEs are non-LTR elements of the L2 and CR1 superfamilies, and SINEs (Venkatesh et al. 2005, 2007, 2014; Chalopin, et al. 2015). Previous findings indicate that SINEs are well represented in the genomes of sharks and rays, suggesting that the TE landscapes in cartilaginous fishes might be more similar to jawless than to bony fishes (Ogiwara et al. 1999, 2002). By contrast, the low relative contribution of TEs other than non-LTRs retrotransposons makes the elephant shark genome an exception among fishes regarding TE diversity. Additionally, the elephant shark genome is one of the slowest-evolving among vertebrates (Venkatesh et al. 2014) and this is reflected by the presence of many old, degenerate TEs.

### Actinopterygii

The numbers and proportions of TEs are extremely variable among genomes of actinopterygian (ray-finned) fishes, especially teleosts, which exhibit the highest number of TE superfamilies among vertebrates (Duvernell et al. 2004; Volff 2005). This is often reflected in genome sizes within the clade. For example, Teleostei include the smallest reported vertebrate genomes in the green spotted pufferfish and the fugu (*Takifugu rubripes*), ~342 and 393 Mb, respectively, which consist of only ~6% of TE-derived DNA (Crollius et al. 2000; Aparicio et al. 2002; Volff et al. 2003). But the genome of another teleost, the zebrafish is TE-rich, with ~55% TE content in a genome of ~1.4 Gb (Howe et al. 2013). In fact, TE abundance appears to be the major determinant of genome size across this group (Chalopin et al. 2015; Gao et al. 2016). Chalopin et al. (2015) showed that actinopterygian genome size may be more heavily dependent on TEs than the larger sarcopterygian genomes (including Tetrapoda). The latter exhibit a greater contribution of low copy number and non-repeated sequences.

An average of 24 TE superfamilies per sequenced genome exemplifies the distinctive TE diversity of actinopterygians. Teleost fishes display the highest diversity, reaching 27 TE superfamilies in the zebrafish genome (Howe et al. 2013). Interestingly, extreme genome size reduction in some teleosts did not result in a decrease in TE diversity, contrasting with the major loss of entire TE superfamilies in other vertebrates that also exhibit small genomes. For example, despite the very low overall abundance of TEs in their small genomes, all major types of TEs have been identified in pufferfishes. Even the

number of retrotransposon superfamilies in the fugu and the green spotted pufferfish surpasses that of sarcopterygian lineages, and is significantly higher than the observed in human and mouse (Croliius et al. 2000; Waterston et al. 2002; Kasahara et al. 2007; Church et al. 2009; Chalopin et al. 2015). Thus, as suggested by Furano et al. (2004), there might be two basic host strategies for dealing with TE accumulation – tolerance of increased diversity coupled with lower copy numbers per family or decreased diversity coupled with tolerance of increased copy numbers.

The dominant types of TEs differ drastically among actinopterygian genomes. DNA transposons are a major component of some teleost genomes. For example, class II elements make up ~60% of the TEs in cichlids, and 39% of the zebrafish genome, while retrotransposons make up less than 12% of each (Howe et al. 2013; Brawand et al. 2014). On the other hand, the fugu and the non-teleost spotted gar (*Lepisosteus oculatus*) genomes are characterized by a predominance of non-LTR retrotransposons (Voff et al. 2003; Braasch et al. 2016). Furthermore, the prevalence and diversity of major types of TEs can vary substantially even between closely related groups. The green spotted pufferfish genome differs slightly from that of the fugu, its close relative, by exhibiting a lower diversity of non-LTR retrotransposons. Whereas the prevalence of LINEs and SINEs characterizes the fugu genome, the relative proportions of DNA transposons, LTR and non-LTR elements are roughly equal in the green spotted pufferfish genome (Voff et al. 2003; Chalopin et al. 2015). So far, no actinopterygian genome seems to be particularly enriched for LTR retrotransposons, although many of them, including currently active elements, have been reported in several species (Poulet et al. 1994; Herniou et al. 1998; Poulter and Butler 1998; Voff et al. 2001; Shen and Steiner 2004; Kambol and Abtholuddin 2008; Gao et al. 2016). Finally, there are many cases of lineage-specific losses of TE superfamilies. For instance, the non-LTR retrotransposon Rex3 is widespread among most teleosts, but is not found in salmonids (Voff et al. 2001).

Rates of TE accumulation are also heterogeneous in ray-finned fishes. Recent accumulation has been shown for many TE families (Böhne et al. 2012; Gao et al. 2016). For example, in the medaka (*Oryzias latipes*, teleost) DNA transposon families Tol1 and Tol2 have ongoing transposition bursts and are considered major sources of genetic variation in natural populations (Koga et al. 2006; Tsutsumi et al. 2006; Koga et al. 2009; Watanabe et al. 2014). Among non-LTR retrotransposons, multiple bursts of amplification have been observed in different fish lineages (Voff et al. 2000, 2001). Such differential rates of TE amplification can result in a high turnover of TE families in teleost genomes. Bursts of amplification of few TE families and the elimination of older insertions through large deletions, ectopic recombination, or high nucleotide substitution rates can result in the prevalence of the most recently active elements and very distinct TE landscapes among the

genomes of some closely related species (Duvernell et al. 2004; Blass et al. 2012; Chalopin et al. 2015). Interestingly, similar patterns are also seen in some reptiles (see below).

### Sarcopterygii

Within sarcopterygians (excluding Tetrapoda), the only genomes analyzed in meaningful ways are the African coelacanth (*Latimeria menandoensis*, Amemiya et al. 2013; Nikaido et al. 2013) and the Australian lungfish (*Neoceratodus forsteri*, Metcalfe et al. 2012). Depending on the study, TEs contribute either 25% or 50% of the total genome size of the coelacanth. For example, Nikaido et al. (2013) found that both class I and II TEs contribute to diversity in roughly equal proportions (23% DNA transposons, 26% retrotransposons) but other studies identify fewer DNA transposons (Amemiya et al. 2013; Chalopin et al. 2015). Numerous lineage-specific insertions have occurred in the two extant coelacanth species (Forconi et al. 2014; Naville et al. 2014). Most of these are non-LTR retrotransposons (CR1 LINEs and LF-SINEs), but some DNA transposons have undergone recent transposition (Naville et al. 2014; Chalopin et al. 2015; Naville et al. 2015). Surprisingly, Harbinger DNA elements, which were thought to be an extinct clade of TEs, show evidence of ongoing accumulation (Smith et al. 2012). Recent amplification of non-LTR retrotransposons is one example of many instances where the slow-evolving coelacanth genome has retained activity of TEs that were exapted as regulatory and coding sequences in other vertebrate lineages (Bejerano et al. 2006; Nishihara et al. 2006).

Lungfishes have the largest vertebrate genomes identified to date (~49–127 Gb) (Gregory 2002) and harbor correspondingly massive numbers of TEs. Their large size and high repetitive content has impeded whole-genome sequencing and assembly efforts. Based on survey sequencing, the TE content of the Australian lungfish (*Neoceratodus forsteri*) is estimated to be ~40%. CR1 and L2 LINEs make up over half of those TEs, and correspond to ~22% of the genome (Metcalfe et al. 2012). Transcriptome analyses of the West African lungfish (*Protopterus annectens*) show a high diversity of transcribed TEs (Biscotti et al. 2016). The most prevalent are LINEs, followed by DNA transposons and SINEs. This transcriptional profile is very similar to the coelacanth (Forconi et al. 2014; Biscotti et al. 2016).

### Amphibia

Very little is known about the diversity and distribution of TEs in amphibians. In part, this is driven by the lack of genomic resources for this group, which are limited to genome drafts of the western clawed frog (*Xenopus tropicalis*) and the Tibetan frog (*Nanorana parkeri*) (Hellsten et al. 2010; Sun, et al. 2015). Around one-third of the clawed frog genome is derived from TEs with roughly three quarters of that being from DNA transposons (Hellsten et al. 2010; Chalopin et al. 2015). In fact

percentage-wise, the clawed frog contains a higher ratio of class II to class I content than any other vertebrate examined to date (Chalopin et al. 2015). By contrast, LTR retrotransposons appear to dominate the genome of the Tibetan frog where these TEs have undergone a rapid expansion in the last 30 my (Sun et al. 2015). This expansion has contributed directly to the larger genome size of the Tibetan frog when compared with the clawed frog.

Due to their massive genomes (14–50 Gb), no genome assembly is available for salamanders, but survey sequencing reveals that between 25% and >47% of the genomes are derived from TEs and that the LTR superfamily Ty3/gypsy dominates the TE landscape (Gregory 2002; Sun, Shepard, et al. 2012; Metcalfe and Casane 2013). The recent assembly of the two smallest chromosomes of the Mexican axolotl (*Ambystoma mexicanum*) recovered similar densities but most of these TEs are yet to be classified (Keinath et al. 2015). In contrast to the clawed frog, salamanders have a higher ratio of LTR retrotransposons to TE content than any other vertebrate investigated to date (Sun, Shepard, et al. 2012). It is not known whether increased TE accumulation rates (Sun, Shepard, et al. 2012; Metcalfe and Casane 2013), decreased rates of DNA loss (Sun, Arriaza, et al. 2012), some combination of both, or other mechanisms are responsible for the genomic gigantism seen in the salamanders. Similar phenomena have been observed in coniferous trees which have similarly massive genome sizes (Nystedt et al. 2013). Regardless, the lack of information on TE dynamics in Anura, Caudata, and Gymnophiona represents a major gap in our knowledge and suggests an area ripe for exploration.

### Squamata

The squamate reptiles likely harbor a diverse array and distribution of TEs similar to fish. However, this clade is like amphibians in that it is poorly represented with regard to high-quality genome assemblies. The first squamate genome assembly was from a lizard, the green anole (*Anolis carolinensis*) with a TE content of approximately 30%. Similar to amphibian and fish genomes, all major groups of elements are present in the anole, but most of them are DNA transposons and LINES (Alföldi et al. 2011). Based on genetic distances, most element families appear to have been, or are, recently active. For example, elements from all five major non-LTR retrotransposon superfamilies are active, but these families are present in relatively low copy number, making the repeat profile more “fish-like” (i.e. high diversity, low copy number) than other amniotes (Novick et al. 2009). In addition to the current and recently active elements, older, more heavily mutated elements appear to have been removed through high rates of DNA loss via ectopic recombination or are undetectable due to high substitution rates (Novick et al. 2009; Tollis and Boissinot 2013).

To date, only two snake genome assemblies have been published. One of these, the cobra (*Ophiophagus hannah*) was not interrogated for TE content (Vonk et al. 2013). The python, however, has been informative, especially when the data were combined with survey sequencing of other snakes. Analyses suggest that despite similar genome sizes, snake species have drastically different TE content but low diversity in the types of TEs present. For example, the copperhead (*Agkistrodon contortrix*) and the Burmese python (*Python molurus bivittatus*), each have a genome size of ~1.4 Gb, but TEs (most of them CR1 LINES) occupy twice the space in the copperhead genome (45%) compared with the python (21%) (Castoe et al. 2011, 2013).

Other squamate genomes are available and have only received minimal annotation efforts with regards to the repetitive portions of their genome. According to this limited information, LINE elements are the dominant TE type in the Asian glass lizard (*Ophisaurus gracilis*) (Song et al. 2015), Japanese gecko (*Gekko japonicus*) (Liu et al. 2015), and bearded dragon (*Pogona vitticeps*) (Georges et al. 2015) genomes. Repeat content ranges from ~40% in the bearded dragon to ~48% in the Japanese gecko. Drawing further conclusions from these three genomes is difficult though, since about half of the identified repeats were unclassified and refined classifications are lacking.

Like some fish, data from squamates suggest the potential for frequent horizontal transfer events. For example, very similar, and, therefore, likely horizontally transferred, SPIN elements are found in at least 17 squamate lineages, some of which diverged more than one hundred million years ago (Gilbert et al. 2011). Furthermore, BovB, a member of the RTE superfamily of LINES, is found across Squamata but its distribution outside of the squamates is disjointed and sparse, including arthropods, monotremes, ruminants, and sea urchin. Based on species phylogenies and BovB distributions, it is likely that BovB is vertically inherited in the reptiles, but has been horizontally transferred from reptilian hosts to other taxa on at least nine different occasions (Walsh et al. 2013).

### Testudines

Turtles have served as a model to study TE evolution for 30 years (Endoh and Okada 1986). The first LINE/SINE partnership was discovered in turtles (Kajikawa et al. 1997; Terai et al. 1998). Yet, in spite of this history, not much is known about the distribution and diversity of TEs in Testudines. The western painted turtle (*Chrysemys picta bellii*), green sea turtle (*Chelonia mydas*), and Chinese soft-shell turtle (*Pelodiscus sinensis*) appear to be intermediate to the squamates and birds with regard to TE content. Around 10% of each genome is derived from TEs (Shaffer et al. 2013; Wang et al. 2013). CR1 LINES are the dominant family in the genomes examined thus far, accounting for a majority of



identified TEs in each (Shaffer et al. 2013) and reflecting much of the ancestral CR1 diversity of amniotes (Suh 2015; Suh et al. 2015). Like amphibians, this clade represents a potential fount of information to gain a better understanding of TE dynamics and impacts.

### Crocodylia

Similar to turtles, crocodylians exhibit low neutral mutation rates and their genomes contain a plethora of recognizable ancient TEs and endogenous viruses (Green et al. 2014; Suh et al. 2014). Comprehensive TE annotations of the genomes of representatives from all three extant families of Crocodylia, namely the saltwater crocodile (*Crocodylus porosus*), gharial (*Gavialis gangeticus*), and American alligator (*Alligator mississippiensis*), revealed that ~37% of each genome is TE-derived. Approximately 95% of these elements belong to families that were active in the common ancestor of the three families (Green et al. 2014). CR1 LINES, the most abundant group of TEs in crocodylians, and other TEs show an overall trend of decreased TE activity and diversity since the crocodylian ancestor. More precisely, crocodylian genomes exhibit a similar diversity of ancestral, amniote CR1 lineages as turtles (Suh 2015; Suh et al. 2015). However, presence/absence analysis of CR1 insertions suggests that members from only one of these lineages were active since the common ancestor of Crocodylia (Suh et al. 2015). While there appears to be some very recent or ongoing CR1 activity in gharial (Suh et al. 2015), the majority of within-crocodylian TE activity is derived from a variety of LTR retrotransposons (superfamilies ERV1, ERV2, ERV4) (Chong et al. 2014) and, to a much lesser extent, two families of Tx1-mobilized SINEs with snRNA-derived heads (Kojima 2015).

### Aves

Among vertebrates, birds are unusual in that they exhibit relatively low copy numbers and a reduced overall diversity of TEs (Hillier et al. 2004; Dalloul et al. 2010; Warren et al. 2010). A typical 1.0–1.3-Gb bird genome harbors between 130,000 and 350,000 TE copies making up only 4.1–9.8% of its size (Hillier et al. 2004; Warren et al. 2010; Poelstra et al. 2014; Zhang et al. 2014). The only clear outlier, the downy woodpecker (*Picoides pubescens*), contains ~700,000 TE copies making up 22.2% of its 1.2-Gb genome assembly (Zhang et al. 2014). The majority of avian TEs belong to the CR1 superfamily (Hillier et al. 2004; Warren et al. 2010; Zhang et al. 2014). Notably, the diversity of CR1 in birds comprises 14 recognized families which emerged from a single CR1 lineage after the bird/crocodylian split, while the rest of the ancient amniote CR1 diversity was lost (Suh 2015; Suh et al. 2015). Analyses of CR1 landscapes and CR1 presence/absence markers suggest that many of these CR1 families were active simultaneously and throughout large parts of avian diversification (Kaiser et al. 2007; Kriegs et al. 2007; St John and

Quinn 2008; Suh et al. 2011; Matzke et al. 2012; Suh et al. 2012); however, evidence for very recent or ongoing CR1 activity is limited to *Tachybaptus ruficollis*, the little grebe (Suh et al. 2012).

LTR retrotransposons constitute the second-largest fraction of TEs in neognaths (chicken + duck and Neoaves) (Zhang et al. 2014), where they have been active throughout their early evolution (Suh et al. 2011a, 2011b, 2015). Although avian LTRs were initially described in the chicken lineage (Hillier et al. 2004; Wicker et al. 2005), there appears to have been more LTR accumulation in Neoaves, especially during their early radiation (Suh et al. 2011, 2015). Among Neoaves, oscine songbirds (e.g., zebra finch, collared flycatcher, American and hooded crow) exhibit increased numbers of young LTR retrotransposons from the superfamilies ERV1, ERV2, and ERV3 (Warren et al. 2010; Cui et al. 2014; Smeds et al. 2015; Vijay et al. in revision). In the collared flycatcher (*Ficedula albicollis*), the recent or ongoing LTR activity coincides with a gradual reduction of CR1 activity and potential lack of ongoing CR1 activity (Smeds et al. 2015). Interestingly, similar trends are visible in the zebra finch (*Taeniopygia guttata*) and the hooded crow (*Corvus cornix*) (Kapusta and Suh 2016). Each of these potential CR1 extinctions occurred very recently and thus each postdate the most recent ancestor of songbirds. This suggests that CR1 extinction and LTR dominance emerged independently in each of these songbird lineages (Kapusta and Suh 2016).

Despite the relative scarcity of TE copies and diversity in birds, more peculiarities are being identified as more genomes become the focus of sequencing and TE research efforts. Most birds are similar to chicken, in that they lack any recent SINE accumulation (Hillier et al. 2004; Zhang et al. 2014). However, CR1-mobilized SINEs have been accumulated in the lineage of zebra finch and related songbirds (Warren et al. 2010; Zhang et al. 2014). On the other hand, the chicken has experienced recent mariner and hAT DNA transposon activity (Hillier et al. 2004; Wicker et al. 2005), which is unusual among birds (Zhang et al. 2014). Finally, some avian lineages (some songbirds, some parrots, hornbills, trogons, hummingbirds, mesites, tinamous) have been infiltrated repeatedly by AviRTE, a newly discovered family of RTE LINES (Suh et al. 2016). This family of RTEs is distantly related to the aforementioned BovB, and the presence of AviRTE in filarial nematodes strongly suggests horizontal transfer between and among birds and nematodes. Notably, SINEs mobilized by AviRTE evolved independently in some songbirds, some parrots, and hornbills (Suh et al. 2016).

### Mammalia

In terms of TE diversity, distribution, and evolution, mammals are the best-studied vertebrates. This is largely due to the higher number of sequenced genomes spanning major lineages within the group (Koeplfli et al. 2015). In addition, there is

substantial information on the impact of TEs for the evolution of genome architecture, even for groups still lacking whole-genome drafts (Wichman et al. 1992; Acosta et al. 2008; Cantrell et al. 2008; Khalil and Driscoll 2010). TEs account for more than half of the size of many mammalian genomes. However, they typically exhibit low subfamily diversity compared to fishes, amphibians, and reptiles, high copy numbers of retrotransposons, and minimal DNA transposon content (Furano et al. 2004). Deviations from this pattern are described below.

### Monotremata

Most of what is known about TEs in monotremes is derived from the platypus (*Ornithorhynchus anatinus*) genome assembly (Warren et al. 2008). Similar to monotreme morphology, the TE landscape exhibits characteristics that are intermediate between reptiles and mammals. Interspersed repeats derived from TEs account for approximately 45% of the platypus genome. The most prevalent TEs, with over 1.5 million copies each, are L2 and the co-mobilized MIR/Mon-1 SINE, both of which went extinct in therians (Metatheria and Eutheria) around 60–100 Mya. Monotremes have a novel SINE-like, small nucleolar RNA-derived retrotransposon that is mobilized by an RTE (sno-RTEs) rather than by L1, the common LINEs of therians (Schmitz et al. 2008; Warren et al. 2008). Interestingly, the patterns of genomic distribution and past accumulation of DNA transposons and LTR retrotransposons differ between monotremes and therians. In the platypus genome, DNA transposons and LTR retrotransposons are particularly underrepresented in genic regions known to undergo imprinting in therians. This suggests that therian-specific expansion and accumulation of these TEs may have promoted the evolution of imprinting, which is a feature present in metatherian and eutherian but not monotreme genomes (Pask et al. 2009).

### Metatheria

The genomes of Metatheria (marsupials) more closely resemble the genomes of eutherian mammals than monotremes. Analyses of the short-tailed opossum (*Monodelphis domestica*), tammar wallaby (*Macropus eugenii*), and Tasmanian devil (*Sarcophilus harrisi*) draft genomes suggest that over half of the typical marsupial genome is derived from TEs (Mikkelsen et al. 2007; Renfree et al. 2011; Nilsson et al. 2012; Nilsson 2016). As with other therians, large fractions of TEs correspond to non-LTR elements, such as LINEs and SINEs, including a high fraction of RTEs that are widely distributed across all marsupial orders (Gentles et al. 2007). By contrast, RTEs, specifically BovB, are restricted to only a few eutherian lineages and these instances are likely due to horizontal transfer events (Walsh et al. 2013). SINE activity is thought to have ceased before the radiation of the Dasyuridae (marsupial mice, quolls, and Tasmanian devils)

~30 Mya, followed by L1 extinction in the lineage of the Tasmanian devil (but see Nilsson, 2016, for further comments). The opossum however, harbors transcriptionally active L1 and SINEs (Gu et al. 2007).

### Eutheria

Most modern eutherians (placental mammals) share a core set of TE characteristics that includes a relatively high TE content but lower TE diversity when compared with non-mammalian and non-avian vertebrates. Additional features characterizing most eutherian genomes are the significant degeneration of ancient vertebrate elements and the underrepresentation and inactivity of DNA transposons (Chalopin et al. 2015). However, many ancient TEs are still visible because they have been exapted as conserved non-coding elements, a large fraction of which were inserted in the common ancestor of eutherians (Mikkelsen et al. 2007; Jurka et al. 2012).

With few exceptions (e.g., see Adelson et al. 2009; Walsh et al. 2013), the most common TE in many eutherian genomes is L1 which, unlike many non-mammalian retrotransposon families, typically exhibits a single lineage of successive subfamilies (Boissinot and Furano 2001; Furano et al. 2004). L1-dependent SINEs recurrently arose de novo, and often make up significant fractions of therian genomes. These SINEs are usually order specific. Approximately 67 SINE families have been described in mammals, 29 of which are eutherian specific (Shimamura et al. 1999; Gogolevsky et al. 2009; Churakov et al. 2010; Kramerov and Vassetzky 2011; Vassetzky and Kramerov 2013). There is variation in this characteristic, however. Some lineages harbor multiple active SINEs (Kass et al. 2000; Kass and Jamison 2007), whereas others exhibit either no evidence of SINE accumulation (Rinehart et al. 2005; Platt and Ray 2012), or only minimal accumulation in the recent past, like the orangutan genome, in which the primate SINE Alu has apparently generated only ~250 insertions over the past 12 million years (Locke et al. 2011).

While most TE activity in eutherian genomes is represented by retrotransposition of L1 and SINEs (Akagi et al. 2008; Chalopin et al. 2015), L1 extinctions (and associated SINE shutdowns) have been reported in some lineages (Casavant et al. 2000; Grahn et al. 2005; Cantrell et al. 2008; Platt and Ray 2012). Interestingly, in muroid rodents, L1 extinction events are correlated with a shift or genomic invasion by other types of elements, ERVs (Cantrell et al. 2005; Erickson et al. 2011). Indeed, mice are atypical mammals in that they have experienced significant LTR retrotransposon accumulation in addition to the typical mammalian LINEs and SINEs (Nellaker et al. 2012). By contrast, most species of eutherians exhibit only modest contributions of ERV and ERV-like elements when compared to other TEs (Bénil et al. 1999; Mager and Stoye 2015).

Significant DNA transposon activity ceased in mammals around 40 Mya (Pace and Feschotte 2007), with a few minor (Zhao et al. 2009; Pagan et al. 2010) and one major exception. Significant accumulation of DNA transposons has been found in a single family of bats, Vespertilionidae (Pritham and Feschotte 2007; Ray et al. 2007, 2008; Pagán et al. 2012; Mitra et al. 2013; Platt et al. 2014; Thomas et al. 2014). Analyses confirm that multiple class II superfamilies have been accumulating in vesper bat genomes in the recent past, but not in other closely related taxa (Platt et al. 2016). This includes the rolling-circle transposons (Helitrons), which comprise over 100,000 copies in the genome of *Myotis lucifugus* (Pritham and Feschotte 2007; Thomas et al. 2014). These observations have generated particular interest in that the massive amount of accumulation by these elements may be associated with the high rates of diversification in this clade via perturbation of regulatory networks and the generation of genomic novelties (Platt et al. 2014; Thomas et al. 2014), lending support to the growing number of hypotheses that relate TE accumulation and species diversity (Zeh et al. 2009; Oliver and Greene 2011; Stapley et al. 2015).

## The Future

As demonstrated above, advances in sequencing technology have been a boon to the field of TE biology. The increased throughput provided by new sequencing technologies has allowed for an unprecedented rate of discovery. The benefit of so many new genome assemblies is obvious. As new genomes from a broad range of taxa become available, we have the opportunity to expand our knowledge of the TE landscapes in each one. In particular, this is demonstrated by the increasing number of manuscripts that detail not one genome but several, a trend that will likely increase. In one such recent example, Zhang et al. (2014) published assemblies for 45 new avian genomes. Although the consortium only performed uncurated RepeatModeler (Smit and Hubley 2008–2015) de novo analysis of the TE content in each genome (the shortcomings of which are discussed above), they revealed some interesting findings, including a substantial number of LINE elements in the woodpecker and the large copy numbers of LTR retrotransposons specific to songbirds. These examples provide evidence of lineage-specific expansions of particular TE groups that could have served to influence the evolution of birds in any number of ways. However, in-depth TE curation of most of these genomes is pending and some ongoing analyses have unearthed fascinating findings (e.g., Suh et al. 2016).

Several other groups are taking advantage of our increased ability to generate sequence data to assemble genomes for a large number of species in particular taxonomic groups. The Broad Institute, for example, recently embarked on an effort to generate genome assemblies for 150 additional mammal species (Johnson J, personal communication). As part of that effort, they have contacted individuals with expertise to

determine which species contain the most scientific value. As these genomes are released; however, the TE community must be ready to identify the repetitive complements in each. The same applies for the B10K project of BGI Shenzhen, which recently announced plans to sequence all 10,500 species of birds (Zhang G, personal communication).

The decrease in sequencing costs has greatly increased the availability of non-model genome assemblies. However, generating an assembly for every genome of interest is still a major undertaking and is not a viable strategy for many laboratories. Fortunately, information on the TE content of a genome is readily available even without generating a de novo assembly. Because most TEs are, by definition, present in multiple copies, their composition and impact can be investigated via survey sequencing. For example, a subfamily of SINE in a genome, even when present at only 1,000 copies (a relatively small number for a SINE), is still present at 1,000 times the single-copy portions. Thus, even when sequencing a genome at a depth of only 0.5X coverage, one would expect to find multiple copies of that SINE (or at least an increased read depth). Multiple studies have illustrated this point in vertebrates. For example, efforts to analyze the genomes of vesper bats revealed the unique dynamics of DNA transposons and Ves SINEs in bats, including substantial difference in SINE accumulation when comparing vesper bats to other families and several lineage-specific subfamilies within vesper bats (Pagán et al. 2012; Ray et al. 2015). Sun, Shepard, et al. (2012) used similar methods to implicate LTR retrotransposons in the evolution of large genome size in plethodontid salamanders and Castoe et al. (2011) identified substantial TE diversity among snake genomes. In contrast to the studies mentioned above which used 454-based sequencing chemistry, Castoe et al. (2011) employed Illumina sequencing technology. While this chemistry generates vast amounts of data, the short read lengths can be a limiting factor in its utility. For example, given the fact that most non-SINE TEs are several hundred nucleotides or longer, obtaining full-length elements using Illumina data is nearly impossible, even when one generates the longest possible reads and overlapping sequencing libraries, or uses recent TE analyses pipelines that assemble survey sequencing reads into longer contigs (e.g., dnaPipeTE; Goubert et al. 2015).

Fortunately, Pacific Biosciences (PacBio) SMRT sequencing (Eid et al. 2009) may serve to alleviate this problem. Single PacBio reads can average over 10,000 nt, more than enough to span all but the largest full-length TE insertions. The high error rate exhibited by PacBio chemistry could present a problem for certain types of analyses. However, the most common way to overcome this weakness is to sequence the region multiple times over, thereby generating enough reads to correct the data via consensus methods. Given the overabundance of TEs in any given genome, multiple insertions will likely be present in the data and, while these multiple insertions cannot be used to correct errors in any given insertion,

they can be used to identify the consensus for any particular family that might be present. Additionally, as costs for long-read sequencing decrease, it will soon be possible to reassess the influence of TE content of structurally complex regions that are often missed in short-read assemblies. To our knowledge, no one has yet used PacBio in this way but it seems a natural extension of the technology.

## Conclusions

At the most basic level of inquiry, the percent of a genome derived from TEs, vertebrate genomes can vary from 6 to 60%. If one takes into account aspects of TE diversity, accumulation histories, and even variation in repeat annotations themselves, it becomes difficult to build a coherent narrative that adequately explains repeat variation across vertebrates. Generally, higher levels of TE diversity correlate with the age of vertebrate lineages; lineages that have existed for longer periods, such as fishes, and deep-branching tetrapods tend to have higher TE diversity than more recent radiations, such as birds and mammals. However, as the number of vertebrate genome assemblies increases, exceptions to this pattern will become more common. Known outliers within each vertebrate lineage include the lungfish with a genome dominated by two types of non-LTR retrotransposons, and the western clawed frog whose TE content is highly biased towards DNA transposons. Woodpeckers contain almost half a million more TE copies than other birds. Among mammals, vespertilionid bats are the sole lineage exhibiting DNA transposon activity. Indeed, our view of what is “normal” for broad lineage such as mammals or birds continues to expand and our understanding of TEs and their role in vertebrate genome evolution benefits greatly from understanding both general trends and outliers. Identification of the contribution of TEs to the uniqueness of each genome will be key to unraveling the impact of genome architecture on organismal evolution.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This work was supported by the National Science Foundation (DEB-1355176, DEB-1020865, MCB-0841821 and MCB-1052500 to D.A.R.). Additional support was provided by College of Arts and Sciences at Texas Tech University. C.G.S.C. was supported by a Postdoctoral scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq-Brazil.

## Literature Cited

Acosta MJ, Marchal JA, Fernández-Espartero CH, Bullejos M, Sánchez A. 2008. Retroelements (LINEs and SINEs) in vole genomes: differential

- distribution in the constitutive heterochromatin. *Chromosome Res.* 16:949–959.
- Adelson DL, Raison JM, Edgar RC. 2009. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. *Proc Natl Acad Sci U S A.* 106:12855–12860.
- Akagi K, Li J, Stephens RM, Volfovsky N, Symer DE. 2008. Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. *Genome Res.* 18:869–880.
- Alföldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477:587–591.
- Amemiya CT, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* 496:311–316.
- Aparicio S, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–1310.
- Bao W, Jurka MG, Kapitonov VV, Jurka J. 2009. New superfamilies of eukaryotic DNA transposons and their internal divisions. *Mol Biol Evol.* 26:983–993.
- Bejerano G, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441:87–90.
- Bénit L, Lallemand J-B, Casella J-F, Philippe H, Heidmann T. 1999. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J Virol.* 73:3301–3308.
- Bergman CM, Bensasson D. 2007. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 104:11340–11345.
- Berthelot C, et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun.* 5:3657.
- Biscotti MA, et al. 2016. The Lungfish Transcriptome: a glimpse into molecular evolution events at the transition from water to land. *Sci Rep.* 6:21571.
- Blass E, Bell M, Boissinot S. 2012. Accumulation and rapid decay of non-LTR retrotransposons in the genome of the three-spine stickleback. *Genome Biol Evol.* 4:687–702.
- Boer JGD, Yazawa R, Davidson WS, Koop BF. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8:422.
- Böhne A, et al. 2012. Zisupton—a novel superfamily of DNA transposable elements recently active in fish. *Mol Biol Evol.* 29:631–645.
- Boissinot S, Furano AV. 2001. Adaptive evolution in LINE-1 retrotransposons. *Mol Biol Evol.* 18:2186–2194.
- Braasch I, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet.* 48:427–437.
- Brawand D, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513:375–381.
- Cantrell MA, et al. 2005. MysTR: an endogenous retrovirus family in mammals that is undergoing recent amplifications to unprecedented copy numbers. *J Virol.* 79:14698–14707.
- Cantrell MA, Scott L, Brown CJ, Martinez AR, Wichman HA. 2008. Loss of LINE-1 activity in the megabats. *Genetics* 178:393–404.
- Carbone L, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature* 513:195–201.
- Casavant NC, et al. 2000. The end of the LINE?: lack of recent L1 activity in a group of South American rodents. *Genetics* 154:1809–1817.
- Casola C, Hucks D, Feschotte C. 2008. Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol Biol Evol.* 25:29–41.
- Castoe TA, et al. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci U S A.* 110:20645–20650.
- Castoe TA, et al. 2011. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol Evol.* 3:641–653.

- Chalopin D, et al. 2014. Evolutionary active transposable elements in the genome of the coelacanth. *J Exp Zool B Mol Dev Evol.* 322:322–333.
- 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.* 7:567–580.
- Chénais B, Caruso A, Hiard S, Casse N. 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* 509:7–15.
- Chong AY, et al. 2014. Evolution and gene capture in ancient endogenous retroviruses—insights from the crocodylian genomes. *Retrovirology* 11:71.
- Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351:1083–1087.
- Churakov G, et al. 2010. Rodent evolution: back to the root. *Mol Biol Evol.* 27:1315–1326.
- Church DM, et al. 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.* 7:e1000112.
- Crollius HR, et al. 2000. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res.* 10:939–949.
- Cui J, et al. 2014. Low frequency of paleoviral infiltration across the avian phylogeny. *Genome Biol.* 15:539.
- Dalloul RA, et al. 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* 8:2241.
- Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. 1990. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* 124:339–355.
- 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:e1002384.
- Deininger PL, Batzer MA. 2002. Mammalian retroelements. *Genome Res.* 12:1455–1465.
- Duvernell DD, Pryor SR, Adams SM. 2004. Teleost fish genomes contain a diverse array of 11 retrotransposon lineages that exhibit a low copy number and high rate of turnover. *J Mol Evol.* 59:298–308.
- Eickbush T, Malik H. 2002. Origins and evolution of retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. *Mobile DNA II*. Washington, DC: ASM Press. p. 1111–1144.
- Eickbush TH, Jamburuthugoda VK. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* 134:221–234.
- Eid J, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138.
- Elliott TA, Gregory TR. 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos Trans R Soc B.* 370:20140331.
- Endoh H, Okada N. 1986. Total DNA transcription in vitro: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. *Proc Natl Acad Sci U S A.* 83:251–255.
- Erickson IK, Cantrell MA, Scott L, Wichman HA. 2011. Retrofitting the genome: L1 extinction follows endogenous retroviral expansion in a group of murid rodents. *J Virol.* 85:12315–12323.
- Farré M, Robinson TJ, Ruiz-Herrera A. 2015. An Integrative Breakage Model of genome architecture, reshuffling and evolution. *Bioessays* 37:479–488.
- Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet.* 41:331.
- Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5:103–107.
- Foot AD, et al. 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet.* 47:272–275.
- Forconi M, et al. 2014. Transcriptional activity of transposable elements in coelacanth. *J Exp Zool B Mol Dev Evol.* 322:379–389.
- Furano AV, Duvernell DD, Boissinot S. 2004. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* 20:9–14.
- Gao B, et al. 2016. The contribution of transposable elements to size variations between four teleost genomes. *Mobile DNA* 7:4.
- Ge R-L, et al. 2013. Draft genome sequence of the Tibetan antelope. *Nat Commun.* 4:1858.
- Gentles AJ, et al. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.* 17:992–1004.
- Georges A, et al. 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *GigaScience* 4(1):11.
- Gilbert C, Hernandez SS, Flores-Benabib J, Smith EN, Feschotte C. 2011. Rampant horizontal transfer of SPIN transposons in squamate reptiles. *Mol Biol Evol.* 29:503–515.
- 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.
- Gilbert C, Waters P, Feschotte C, Schaack S. 2013. Horizontal transfer of OC1 transposons in the Tasmanian devil. *BMC Genomics* 14:134.
- Gogolevsky KP, Vassetzky NS, Kramerov DA. 2009. 5S rRNA-derived and tRNA-derived SINEs in fruit bats. *Genomics* 93:494–500.
- Gordon D, et al. 2016. Long-read sequence assembly of the gorilla genome. *Science* 352: aae0344.
- Goubert C, et al. 2015. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biol Evol.* 7:1192–1205.
- Grabundzija I, et al. 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat Commun.* 7:10716.
- 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.* 110:407–415.
- Gray YH. 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet.* 16:461–468.
- Green RE, et al. 2014. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346:1254449.
- Gregory TR. 2002. Genome size and developmental complexity. *Genetica* 115:131–146.
- Gu W, et al. 2007. SINEs, evolution and genome structure in the opossum. *Gene* 396:46–58.
- Hartl D, Lozovskaya E, Lawrence J. 1992. Nonautonomous transposable elements in prokaryotes and eukaryotes. *Genetica* 86:47–53.
- Hedges D, Deininger P. 2007. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res/Fundam Mol Mech Mutagenesis* 616:46–59.
- Hedges DJ, et al. 2004. Differential Alu mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* 14:1068–1075.
- Hellsten U, et al. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328:633–636.
- Herniou E, et al. 1998. Retroviral diversity and distribution in vertebrates. *J Virol.* 72:5955–5966.
- Hillier LW, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
- Hoen DR, et al. 2015. A call for benchmarking transposable element annotation methods. *Mobile DNA* 6(1):
- Howe K, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496:498–503.

- Jurka J, Bao W, Kojima KK. 2011. Families of transposable elements, population structure and the origin of species. *Biol Direct.* 6(1):16.
- Jurka J, Bao W, Kojima KK, Kohany O, Yurka MG. 2012. Distinct groups of repetitive families preserved in mammals correspond to different periods of regulatory innovations in vertebrates. *Biol Direct.* 7:36.
- Kaiser VB, van Tuinen M, Ellegren H. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol Biol Evol.* 24:338–347.
- Kajikawa M, Ohshima K, Okada N. 1997. Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. *Mol Biol Evol.* 14:1206–1217.
- Kajikawa M, Okada N. 2002. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 111:433–444.
- Kambol R, Abtholuddin MF. 2008. Genome structure and characterisation of an endogenous retrovirus from the zebrafish genome project database. New York: Springer.
- Kapitonov VV, Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* 3:998.
- Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A.* 98:8714–8719.
- Kapitonov VV, Jurka J. 2006. Self-synthesizing DNA transposons in eukaryotes. *Proc Natl Acad Sci U S A.* 103:4540–4545.
- Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet.* 9:411–412.
- Kapusta A, Suh A. 2016. Evolution of bird genomes—a transposon's-eye view. *Ann N Y Acad Sci.* Advance Access published December 20, 2016, doi:10.1111/nyas.13295.
- Kasahara M, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719.
- Kass DH, Jamison N. 2007. Identification of an active ID-like group of SINEs in the mouse. *Genomics* 90:416–420.
- Kass DH, Raynor ME, Williams TM. 2000. Evolutionary history of B1 retrotransposons in the genus *Mus*. *J Mol Evol.* 51:256–264.
- Keinath MC, et al. 2015. Initial characterization of the large genome of the salamander *Ambystoma mexicanum* using shotgun and laser capture chromosome sequencing. *Sci Rep.* 5:1–13.
- Khalil AM, Driscoll DJ. 2010. Epigenetic regulation of pericentromeric heterochromatin during mammalian meiosis. *Cytogenet Genome Res.* 129:280–289.
- Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49–63.
- Kim EB, et al. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 479:223–227.
- Koepfli K-P, Paten B, O'Brien SJ. 2015. The genome 10K project: a way forward. *Annu Rev Anim Biosci.* 3:57–111.
- Koga A, Iida A, Hori H, Shimada A, Shima A. 2006. Vertebrate DNA transposon as a natural mutator: the medaka fish Tol2 element contributes to genetic variation without recognizable traces. *Mol Biol Evol.* 23:1414–1419.
- Koga A, Wakamatsu Y, Sakaizumi M, Hamaguchi S, Shimada A. 2009. Distribution of complete and defective copies of the Tol1 transposable element in natural populations of the medaka fish *Oryzias latipes*. *Genes Genet Syst.* 84:345–352.
- Kojima KK. 2015. A new class of SINEs with snRNA gene-derived heads. *Genome Biol Evol.* 7:1702–1712.
- Kokošar J, Kordiš D. 2013. Genesis and regulatory wiring of retroelement-derived domesticated genes: a phylogenomic perspective. *Mol Biol Evol.* 30:1015–1031. mst014.
- Kordis D, Gubensek F. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc Natl Acad Sci U S A.* 95:10704–10709.
- Kramerov D, Vassetzky N. 2011. Origin and evolution of SINEs in eukaryotic genomes. *Heredity (Edinb.)* 107:487–495.
- Kriegs JO, et al. 2007. Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol Biol.* 7:190.
- Krupovic M, Koonin EV. 2015. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat Rev Microbiol.* 13:105–115.
- Kuraku S, Qiu H, Meyer A. 2012. Horizontal transfers of Tc1 elements between teleost fishes and their vertebrate parasites, lampreys. *Genome Biol Evol.* 4:929–936.
- Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Lee H-E, Eo J, Kim H-S. 2015. Composition and evolutionary importance of transposable elements in humans and primates. *Genes Genomics* 37:135–140.
- Lerat E. 2010. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity (Edinb.)* 104:520–533.
- Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet.* 12:615–627.
- Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen F-M. 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75:1083–1093.
- Lim JK, Simmons MJ. 1994. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *Bioessays* 16:269–275.
- Lindblad-Toh K, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478:476–482.
- Liu Y, et al. 2015. Gekko japonicus genome reveals evolution of adhesive toe pads and tail regeneration. *Nat Commun.* 6:1–11.
- Locke DP, et al. 2011. Comparative and demographic analysis of orangutan genomes. *Nature* 469:529–533.
- Louis A, Nguyen NTT, Muffato M, Crollius HR. 2015. Genomic update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics. *Nucleic Acids Res.* 43:D682–D689.
- Lowe CB, Bejerano G, Haussler D. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci U S A.* 104:8005–8010.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72:595–605.
- Lynch VJ. 2016. A copy-and-paste gene regulatory network. *Science* 351:1029–1030.
- 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.* 43:1154–1159.
- Lynch VJ, et al. 2015. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* 10:551–561.
- Mager DL, Stoye JP. 2015. Mammalian endogenous retroviruses. *Microbiol Spectrum* 3:MDNA3-M0009.
- Matzke A, et al. 2012. Retroposon insertion patterns of neoavian birds: strong evidence for an extensive incomplete lineage sorting era. *Mol Biol Evol.* 29:1497–1501.
- McGaugh SE, et al. 2015. Rapid molecular evolution across amniotes of the IIS/TOR network. *Proc Natl Acad Sci U S A.* 112:7055–7060.
- Metcalfe CJ, Casane D. 2013. Accommodating the load: the transposable element content of very large genomes. *Mob Genet Elements* 3:e24775.
- Metcalfe CJ, Filée J, Germon I, Joss J, Casane D. 2012. Evolution of the Australian lungfish (*Neoceratodus forsteri*) genome: a major role for CR1 and L2 LINE elements. *Mol Biol Evol.* 29:3529–3539.
- Michalak P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity (Edinb.)* 102:45–50.

- Mikkelsen TS, et al. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 437:69–87.
- Mikkelsen TS, et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167–177.
- Mills RE, et al. 2006. Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet*. 78:671–679.
- Mitra R, et al. 2013. Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc Natl Acad Sci U S A*. 110:234–239.
- Naville M, Chalopin D, Casane D, Laurenti P, Volff J-N. 2015. The coelacanth: can a “living fossil” have active transposable elements in its genome? *Mob Genet Elements* 5:55–59.
- Naville M, Chalopin D, Volff J-N. 2014. Interspecies insertion polymorphism analysis reveals recent activity of transposable elements in extant coelacanths. *PLoS One* 9:e114382.
- Nellaker C, et al. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol*. 13:R45.
- Nikaido M, et al. 2013. Coelacanth genomes reveal signatures for evolutionary transition from water to land. *Genome Res*. 23:1740–1748.
- Nilsson MA. 2016. The devil is in the details: transposable element analysis of the Tasmanian devil genome. *Mob Genet Elements* 6:1–7.
- Nilsson MA, Janke A, Murchison EP, Ning Z, Hallström BM. 2012. Expansion of CORE-SINES in the genome of the Tasmanian devil. *BMC Genomics* 13:172.
- Nishihara H, Smit AF, Okada N. 2006. Functional noncoding sequences derived from SINES in the mammalian genome. *Genome Res*. 16:864–874.
- Noonan JP, McCallion AS. 2010. Genomics of long-range regulatory elements. *Annu Rev Genomics Hum Genet*. 11:1–23.
- Novick PA, Basta H, Floumanhaft M, McClure MA, Boissinot S. 2009. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. *Mol Biol Evol*. 26:1811–1822.
- Nystedt B, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:579–584.
- Ogiwara I, Miya M, Ohshima K, Okada N. 1999. Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs. *Mol Biol Evol*. 16:1238–1250.
- Ogiwara I, Miya M, Ohshima K, Okada N. 2002. V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Res*. 12:316–324.
- Ohshima K, Okada N. 2005. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet Genome Res*. 110:475–490.
- Oliver KR, Greene WK. 2011. Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. *Mob DNA* 2:8
- Oliver KR, Greene WK. 2012. Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-Thrust hypothesis. *Ecol Evol*. 2:2912–2933.
- Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. *Bioessays* 31:703–714.
- Orgel LE, Crick FH. 1980. Selfish DNA: the ultimate parasite. *Nature* 284:604–607.
- Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res*. 17:422–432.
- Pagán HJ, et al. 2012. Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. *Genome Biol Evol*. 4:575–585.
- Pagan HJ, Smith JD, Hubley RM, Ray DA. 2010. PiggyBac on a primate genome: novel elements, recent activity and horizontal transfer. *Genome Biol Evol*. 2:293–303.
- Pask AJ, et al. 2009. Analysis of the platypus genome suggests a transposon origin for mammalian imprinting. *Genome Biol*. 10:R1.
- Piégu B, Bire S, Arensburger P, Bigot Y. 2015. A survey of transposable element classification systems—a call for a fundamental update to meet the challenge of their diversity and complexity. *Mol Phylogenet Evol*. 86:90–109.
- Platt RN, Blanco-Berdugo L, Ray DA. 2016. Accurate transposable element annotation is vital when analyzing new genome assemblies. *Genome Biol Evol*. eww009.
- Platt RN, et al. 2014. Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. *Mol Biol Evol*. 31:1536–1545.
- Platt RN, Ray DA. 2012. A non-LTR retroelement extinction in *Spermophilus tridecemlineatus*. *Gene* 500:47–53.
- Platt RN, Mangum SF, Ray DA. 2016. Pinpointing the vesper bat transposon revolution using the *Miniopterus natalensis* genome. *Mob DNA* 7:12.
- Poelstra JW, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410–1414.
- Pop M. 2009. Genome assembly reborn: recent computational challenges. *Brief Bioinform*. 10:354–366.
- Poulet FM, Bowser PR, Casey JW. 1994. Retroviruses of fish, reptiles, and molluscs. In: Levy J, editor. *The retroviridae*. New York (NY): Springer. p. 1–38.
- Poulter R, Butler M. 1998. A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*. *Gene* 215:241–249.
- 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.
- Pritham EJ, Putliwala T, Feschotte C. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* 390:3–17.
- Ray DA, Batzer MA. 2005. Tracking Alu evolution in New World primates. *BMC Evol Biol*. 5:51.
- Ray DA, et al. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res*. 18:717–728.
- Ray DA, et al. 2015. Differential SINE evolution in vesper and non-vesper bats. *Mobile DNA* 6:1–10.
- Ray DA, Pagan HJ, Thompson ML, Stevens RD. 2007. Bats with hATs: evidence for recent DNA transposon activity in genus *Myotis*. *Mol Biol Evol*. 24:632–639.
- Rebollo R, Romanish MT, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet*. 46:21–42.
- Renfree MB, et al. 2011. Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol*. 12:R81.
- Rinehart T, Grahn R, Wichman H. 2005. SINE extinction preceded LINE extinction in sigmodontine rodents: implications for retrotranspositional dynamics and mechanisms. *Cytogenet Genome Res*. 110:416–425.
- Sasaki T, et al. 2008. Possible involvement of SINEs in mammalian-specific brain formation. *Proc Natl Acad Sci U S A*. 105:4220–4225.
- Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol*. 25:537–546.
- Schmidt D, et al. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148:335–348.
- Schmitz J, et al. 2008. Retroposed SNOfall—a mammalian-wide comparison of platypus snoRNAs. *Genome Res*. 18:1005–1010.
- Shaffer HB, et al. 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol*. 14:R28.

- Shen C-H, Steiner LA. 2004. Genome structure and thymic expression of an endogenous retrovirus in zebrafish. *J Virol.* 78:899–911.
- 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.* 16:1046–1060.
- Siguier P, Goubeyre E, Varani A, Ton-Hoang B, Chandler M. 2015. Everyman's guide to bacterial insertion sequences. *Microbiol Spectrum* 3: MDNA3-0030-2014.
- Silva JC, Loreto EL, Clark JB. 2004. Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol.* 6:57–71.
- Smeds L, et al. 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nat Commun.* 6: 7330
- Smit A, Hubley R. 2008–2015. RepeatModeler. Seattle (WA): Institute for Systems Biology.
- Smith JJ, Antonacci F, Eichler EE, Amemiya CT. 2009. Programmed loss of millions of base pairs from a vertebrate genome. *Proc Natl Acad Sci U S A.* 106:11212–11217.
- Smith JJ, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet.* 45:415–421.
- Smith JJ, Sumiyama K, Amemiya CT. 2012. A living fossil in the genome of a living fossil: Harbinger transposons in the coelacanth genome. *Mol Biol Evol.* 29:985–993.
- Song B, et al. 2015. A genome draft of the legless anguid lizard, *Ophisaurus gracilis*. *DNA* 56:3.19.
- St John J, Quinn TW. 2008. Identification of novel CR1 subfamilies in an avian order with recently active elements. *Mol Phylogenet Evol.* 49:1008–1014.
- Stapley J, Santure AW, Dennis SR. 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol Ecol.* 24:2241–2252.
- Suh A. 2015. The specific requirements for CR1 retrotransposition explain the scarcity of retrogenes in birds. *J Mol Evol.* 81:18–20.
- Suh A, et al. 2015. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol Evol.* 7:205–217.
- Suh A, Kriegs JO, Brosius J, Schmitz J. 2011. Retroposon insertions and the chronology of avian sex chromosome evolution. *Mol Biol Evol.* 28:2993–2997.
- Suh A, Kriegs JO, Donnellan S, Brosius J, Schmitz J. 2012. A universal method for the study of CR1 retroposons in nonmodel bird genomes. *Mol Biol Evol.* 29:2899–2903.
- Suh A, et al. 2011. Mesozoic retrotransposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun.* 2:443.
- Suh A, Smeds L, Ellegren H. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol.* 13:e1002224.
- Suh A, et al. 2014. Early Mesozoic coexistence of amniotes and Hepadnaviridae. *PLoS Genet.* 10:e1004559.
- Suh A, et al. 2016. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *Nat Commun.* 7:11396.
- Sun C, Arriaza JRL, Mueller RL. 2012. Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol Evol.* 4:1340–1348.
- Sun C, et al. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol Evol.* 4:168–183.
- Sun Y-B, et al. 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc Natl Acad Sci U S A.* 112:E1257–E1262.
- Terai Y, Takahashi K, Okada N. 1998. SINE cousins: the 3'-end tails of the two oldest and distantly related families of SINEs are descended from the 3'ends of LINES with the same genealogical origin. *Mol Biol Evol.* 15:1460–1471.
- Thomas J, Phillips CD, Baker RJ, Pritham EJ. 2014. Rolling-circle transposons catalyze genomic innovation in a mammalian lineage. *Genome Biol Evol.* 6:2595–2610.
- Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol Evol.* 2:656–664.
- Tollis M, Boissinot S. 2013. Lizards and LINES: selection and demography affect the fate of L1 retrotransposons in the genome of the green anole (*Anolis carolinensis*). *Genome Biol Evol.* 5:1754–1768.
- Tsutsumi M, et al. 2006. Color reversion of the albino medaka fish associated with spontaneous somatic excision of the Tol-1 transposable element from the tyrosinase gene. *Pigment Cell Res.* 19:243–247.
- 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.* 19:530–536.
- Vassetzky NS, Kramerov DA. 2013. SINEBase: a database and tool for SINE analysis. *Nucleic Acids Res.* 41:D83–D89.
- Venkatesh B, et al. 2007. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol.* 5:e101.
- Venkatesh B, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505:174–179.
- Venkatesh B, Tay A, Dandona N, Patil JG, Brenner S. 2005. A compact cartilaginous fish model genome. *Curr Biol.* 15:R82–R83.
- Ventura M, et al. 2011. Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee. *Genome Res.* 21:1640–1649.
- Vijay N, et al. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat Commun.* 7:13195.
- Volff J-N, Bouneau L, Ozouf-Costaz C, Fischer C. 2003. Diversity of retrotransposable elements in compact pufferfish genomes. *Trends Genet.* 19:674–678.
- Volff J-N, et al. 2001. Jule from the fish *Xiphophorus* is the first complete vertebrate Ty3/Gypsy retrotransposon from the Mag family. *Mol Biol Evol.* 18:101–111.
- Volff J-N, Körting C, Meyer A, Scharl M. 2001. Evolution and discontinuous distribution of Rex3 retrotransposons in fish. *Mol Biol Evol.* 18:427–431.
- Volff J-N, Körting C, Scharl M. 2000. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. *Mol Biol Evol.* 17:1673–1684.
- Volff J. 2005. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb.)* 94:280–294.
- Vonk FJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci U S A.* 110:20651–20656.
- Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL. 2013. Widespread horizontal transfer of retrotransposons. *Proc Natl Acad Sci U S A.* 110:1012–1016.
- Wang H, et al. 2005. SVA elements: a Hominid-specific Retroposon Family. *J Mol Biol.* 354:994–1007.
- Wang J, Davis RE. 2014. Programmed DNA elimination in multicellular organisms. *Curr Opin Genet Dev.* 27:26–34.
- Wang Z, et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet.* 45:701–706.
- Warren WC, et al. 2010. The genome of a songbird. *Nature* 464:757–762.
- Warren WC, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–183.
- Watanabe K, et al. 2014. Spontaneous germline excision of Tol1, a DNA-based transposable element naturally occurring in the medaka fish genome. *Genome* 57:193–199.



- Waterston RH, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.
- Wheeler TJ, Eddy SR. 2013. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29:2487–2489.
- Wichman HA, Van Den Bussche RA, Hamilton MJ, Baker RJ. 1992. Transposable elements and the evolution of genome organization in mammals. *Genetica* 86:287–293.
- Wicker T, et al. 2005. The repetitive landscape of the chicken genome. *Genome Res.* 15:126–136.
- Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 8:973–982.
- Zeh DW, Zeh JA, Ishida Y. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays* 31:715–726.
- Zhang G, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.
- Zhang H-H, Feschotte C, Han M-J, Zhang Z. 2014. Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals. *Genome Biol Evol.* 6:1375–1386.
- Zhao F, Qi J, Schuster SC. 2009. Tracking the past: interspersed repeats in an extinct Afrotherian mammal, *Mammuthus primigenius*. *Genome Res.* 19:1384–1392.

**Associate editor:** Kateryna Makova