

# The Mobilome of Reptiles: Evolution, Structure, and Function

Stéphane Boissinot<sup>a</sup> Yann Bourgeois<sup>a</sup> Joseph D. Manthey<sup>b</sup>  
Robert P. Ruggiero<sup>a</sup>

<sup>a</sup>New York University Abu Dhabi, Abu Dhabi, UAE; <sup>b</sup>Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA

## Keywords

Birds · Reptiles · Retrotransposons · Transposons

## Abstract

Transposable elements (TE) constitute one of the most variable genomic features among vertebrates, impacting genome size, structure, and composition. Despite their important role in shaping genomic diversity, they have mostly been studied in mammals, which display one of the least diverse genomes in terms of TE diversity. Recent new resources in reptilian genomics have opened a broader perspective about TE evolution in amniotes. We discuss these recent results by showing that TE diversity is high in reptiles, particularly in squamates, with strong heterogeneity in the number of TE classes retained in each lineage, even at short evolutionary scales. More research is needed to uncover the exact mechanisms that regulate TE proliferation in reptiles and to what extent these selfish elements can play a role in local adaptation or in the emergence of barriers to gene flow.

© 2019 S. Karger AG, Basel

Vertebrate genomes differ considerably in size, structure, and composition, although the number of protein-coding genes is similar across vertebrate lineages. One of the genomic features that show the most variation among

vertebrates is the abundance and diversity of transposable elements (TEs) [Tollis and Boissinot, 2012; Sotero-Caio et al., 2017]. TE is an umbrella term that covers a large diversity of mobile DNA sequences that have the ability to replicate and multiply in the genome of their host. TEs have had a profound impact on the evolution of genomic size and structure in vertebrates. The abundance of TEs is indeed one of the main determinants of haploid genome size variation [Kidwell, 2002], and the differential abundance of TEs across genomes contributes indirectly to other genomic features such as regional variations in base composition. In addition, TEs can be a source of genetic novelties that are advantageous to the host [Warren et al., 2015], but TEs can also cause genetic defects either directly (e.g., by inserting into genes) or indirectly (e.g., by mediating ectopic recombination events) and impose a substantial fitness cost on their host [Pasyukova et al., 2004; Boissinot et al., 2006; Ruggiero et al., 2017].

Although it was known since the 1960s that a large fraction of all vertebrate genomes consist of repetitive and mobile sequences [Britten and Kohne, 1968], our knowledge of TE diversity has been heavily biased towards mammals since then, and this situation has remained until very recently. With the exception of the fugu [Aparicio et al., 2002], which was chosen because of its inordinately small genome, the first sequenced vertebrate genomes were mammals including human and mouse [Lander et

al., 2001; Mouse Genome Sequencing Consortium et al., 2002], soon followed by a number of model species and species of economic interest, including the chicken genome in 2004 [International Chicken Genome Sequencing Consortium, 2004]. It took 10 more years after the sequencing of the first mammalian genome, for the first nonavian reptile genome, the green anole *Anolis carolinensis*, to become available [Alföldi et al., 2011]. Even today the number of mammalian genomes publicly available far exceeds the number of reptile genomes, despite the fact that there are about 4 times as many reptile species as there are mammalian species. The current sampling of amphibians is even less representative with only 3 genomes available for a group that includes approximately 7,000 species. From the point of view of TE's abundance and diversity, mammals have extremely derived genomes and, among all vertebrates, the genome of mammals is probably the most different from the ancestral amniote genomes [Tollis and Boissinot, 2012; Sotero-Caio et al., 2017]. Mammalian genomes are dominated by a single type of TE, called LINE-1 or L1, which has evolved as a single lineage since the origin of placental mammals. This extremely low TE diversity is in fact unique among vertebrates, and the investigation of fish, amphibian, and reptile genomes revealed a remarkable diversity of mobile elements [Chalopin et al., 2015], which seems to be the rule in vertebrates and most likely represent the ancestral state of vertebrates in general, and amniotes in particular. It is thus crucial to expand our understanding of TE evolutionary dynamics outside of mammals and in particular in reptiles, the sister group of mammals, to understand the evolutionary forces that have affected genomic structure and function, to reconstruct the repetitive landscape of the ancestral amniote genome, and finally to decipher the mechanisms responsible for the evolution of genome size.

It has been 10 years since the remarkable review of Kordis in *Cytogenetics and Genome Research* [Kordis, 2009], to our knowledge the only review solely dedicated to TEs in reptiles. When this review was published, a single bird and no nonavian reptile genome had been fully sequenced. Since then, many reptile genomes, representative of the main extant lineages, have been sequenced, and our understanding of TE evolution in reptiles has considerably improved. Here, we review the recent literature on reptile TEs with a focus on the contribution of complete genome sequences, and in particular the contribution of the *A. carolinensis* genome, the first nonavian reptile genome to have been sequenced [Alföldi et al., 2011]. We will first introduce the mobilome (the fraction of a genome made

of mobile sequences), and then we will describe the patterns of abundance and diversity of TEs in the main groups of reptiles. In the third section of the review, we will discuss the impact TEs have had on reptile genomes in terms of cost and benefit and how, in the long term, TEs have affected genome size evolution. We will conclude by discussing the current limits to our knowledge, and we will propose research directions for the future.

## What Constitutes the Mobilome?

### *Classification of Transposable Elements*

Typically, TEs are classified into 2 groups based on their mode of transposition: type I, which includes retrotransposons and endogenous retroviruses, and type II, which includes the DNA transposons. Type I elements have an RNA intermediate in their life cycle. Following transcription of the genomic insert, the RNA is reversed transcribed into DNA, by an enzyme called reverse transcriptase. Within type I, there are 2 classes, based on the presence or absence of long terminal repeats (LTR). Transposition in LTR-retrotransposons is similar to that of retroviruses, where reverse transcription occurs in the cytoplasm and the cDNA is subsequently inserted in the host genome. The most abundant LTR-retrotransposons in reptile genomes are the *Gypsy*, *Copia*, and *DIRS* elements, which have all been detected in *A. carolinensis* [Alföldi et al., 2011]. The non-LTR retrotransposons, which include the LINEs and the *Penelope* elements, transpose by a process called target-primed reverse transcription, in which reverse transcription takes place at the site of insertion [Luan et al., 1993; Cost et al., 2002; Ichihyanagi and Okada, 2008]. The LINE retrotransposons constitute an ancient and diverse group of elements that are widespread across the eukaryote tree of life, and are classified into 25 major clades [Kapitonov et al., 2009]. The *L1*, *L2*, *CR1*, and *RTE* clades are the most significant from the point of view of reptile genomic evolution and can all be simultaneously active in the same genomes as shown in *A. carolinensis* [Novick et al., 2009]. Despite a strong *cis* preference, the biochemical machinery encoded by non-LTR retrotransposons can act on other transcripts and is responsible for the amplification of nonautonomous SINEs [Kajikawa and Okada, 2002; Dewanieux et al., 2003]. SINEs are short elements derived either from tRNA or 7SL RNA that have the ability to recruit the reverse transcriptase of LINEs, sometimes, but not always, because of similarity between the 3' end of the SINE with the 3' end of its LINE partner [Ohshima et al.,

1996]. SINEs can outcompete their LINE relatives and amplify to extremely large numbers. This process is exemplified in the lizard *A. carolinensis*, which contains 2 LINE/SINE pairs, the *Sauria SINE* mobilized by the *Bov-B RTE* element and the *SINE2* element mobilized by *L2* [Piskurek et al., 2009].

Type II elements have in common that they do not use an RNA intermediate during their transposition, but otherwise they constitute a very heterogeneous group of elements, comprised of cut-and-paste transposons, *helitrons* and *polintons*. The cut-and-paste DNA transposons constitute the most abundant and diverse category of DNA transposons and include 15 superfamilies, with *Tc1/Mariner* and *hAT* elements being the most significant in vertebrate genomic evolution [Feschotte and Pritham, 2007]. These elements have a simple structure with a single open reading frame encoding a transposase and flanked by inverted terminal repeats. Transposition requires recognition of the terminal repeats by the transposase, which then cuts the DNA and inserts it elsewhere. Although DNA transposons move through a cut-and-paste process (vs. copy-and-paste process in retrotransposons), they can increase in copy number if transposition takes place during replication or if repair of an empty site occurs by homologous recombination with a chromosome still containing the insertion [Feschotte and Pritham, 2007]. Type II autonomous elements can also mediate the transposition of nonautonomous elements that are typically, but not always, deleted versions of autonomous copies that have retained the terminal repeats [Hartl et al., 1992]. This is exemplified in *A. carolinensis*, where 4 different types of type II transposons (*hAT*, *Tc1/Mariner*, *helitron*, and *Chapaev*) have mobilized 56 identifiable nonautonomous families of elements that are several orders of magnitude more abundant than their autonomous counterparts [Novick et al., 2011].

#### *Modes of Transmission: Vertical versus Horizontal*

As obligatory genomic parasites, the fate of TEs is intrinsically related to the reproduction of the host. A relationship between transposition and reproduction is implicit since the spread of TEs and the increase in copy number requires transposition to occur in the germline or in early embryogenesis. In non-LTR retrotransposons, the phylogeny of the elements tends to recapitulate the phylogeny of the host, which is consistent with a strict model of vertical transmission [Malik et al., 1999; Kordis et al., 2006; Waters et al., 2007; Boissinot and Sookdeo, 2016]. This appears to be the case in the *L1* and *CR1* clades of LINEs, which have persisted and amplified in reptile

genomes since the origin of amniotes, with no evidence of lateral transfer [Suh et al., 2014]. However, another clade of LINEs called *RTE* seems to be prone to lateral transfer [Zupunski et al., 2001]. The first time a lateral transfer of *RTE* was detected it was between reptiles and ruminant bovines; consequently, this element was termed *Bov-B* [Kordis and Gubensek, 1998]. This discovery originally raised some doubts [Malik et al., 1999] but the evidence was quite strong, and since then, multiple cases of *RTE* transfer have been identified in mammals and in reptiles [Gentles et al., 2007; Walsh et al., 2013; Suh et al., 2016].

Since the 1970s, it was known that type II transposons have the ability to jump from host to host in insects [Bartolomé et al., 2009]. Because the germline is sequestered in amniotes, it was thought unlikely that amniote genomes could be colonized de novo by DNA transposons, although the remnants of past type II transposon activity had been detected in human [Pace and Feschotte, 2007]. It thus came as a surprise that a type of DNA transposon, named space invader elements (*SPIN*), has repeatedly colonized vertebrate genomes, in particular reptilian genomes [Pace et al., 2008]. It was subsequently found that the independent horizontal transfer of *SPIN* transposons occurred on more than 12 occasions over the past 50 million years in squamates [Gilbert et al., 2012]. Soon after this discovery was made, several other instances of horizontal transfer were identified in *A. carolinensis* involving other types of transposons [Novick et al., 2010]. This process is not limited to *A. carolinensis*, and it appears that lateral transfer of DNA transposons is quite common in reptiles and that different categories of type II elements have been transferred, including *hAT*, *Mariner*, and *helitrons* [Schaack et al., 2010; Thomas et al., 2010]. Typically, however, invading DNA transposons do not establish persistent active populations of progenitors but instead are short lived. They usually amplify soon after the transfer to the new host, sometimes to extremely large numbers, but rapidly become extinct.

The exact mechanism of TE transfer to the germline is unknown at this point, but recent data suggest a role of parasites and pathogens in TE transmission [Gilbert et al., 2010; Suh et al., 2016]. Indeed, it was shown that some parasites and their host share TEs, suggesting that this type of intimate interaction can facilitate transmission and that parasites may constitute a vehicle for laterally transferred TEs. However, this is only one piece of the puzzle because for the lateral transfer to be successful, the TEs must colonize the germline, either in the gonads, in the environment, or during early embryogenesis, which

poses a problem since amniotes do not release their gametes in the environment and the embryo develops in the amniotic egg where it is protected. Thus, an additional process must exist to move the TE to the germline. It was proposed that viruses, and in particular poxviruses, which have a broad host range, could play a role in this process [Piskurek and Okada, 2007]. The discovery of a snake SINE in a poxvirus known to infect rodents supports this hypothesis [Piskurek and Okada, 2007], but clearly, more research is needed to fully understand the mechanism of lateral transfer in amniotes.

#### *The Evolution of Transposable Elements in Reptiles*

The long-term evolution of TEs within their hosts' genomes is best studied in non-LTR retrotransposons because these elements are almost always transmitted vertically. Such analyses have revealed some striking differences among vertebrate lineages and among clades of non-LTR retrotransposons within the same genome. In mammals, TE diversity is generally limited to a single active clade of non-LTR retrotransposons, namely *L1*. In most species analyzed, *L1* evolves as a single lineage of family, where one family is active at a time, until it is replaced by a novel family that achieves replicative supremacy concomitantly with the extinction of its predecessor [Smit et al., 1995; Boissinot et al., 2000; Khan et al., 2006; Sookdeo et al., 2013]. Consequently, most mammalian genomes contain a single active family of *L1* element at a time. In contrast, *L1* in *A. carolinensis* is represented by at least 20 concurrently active families, which diverged from each other before the split between mammals and reptiles [Novick et al., 2009; Boissinot and Sookdeo, 2016]. This situation is comparable to that of fish and frog genomes, where *L1* is represented by multiple active families, thus indicating that the reduction in diversity of *L1* is a mammalian specific feature [Duvernell et al., 2004; Furano et al., 2004; Boissinot and Sookdeo, 2016; Ivancevic et al., 2016]. This pattern is not limited to *L1* since the *L2* clade is represented in anole by 17 simultaneously active families, although *L2* families are not as divergent from each other as *L1* families are [Novick et al., 2009]. However, other clades harbor a much lower level of diversity. For example, the *R4* clade is represented by a single family and the *CR1* clade by 4 very similar subsets [Novick et al., 2009]. The relatively low diversity of *CR1* in anole is also found in birds, crocodiles, and turtles, although the common ancestor of all amniotes contained 7 distinct lineages [Suh et al., 2018].

Ultimately, the evolution of TEs is driven and constrained by their environment, which is the host genome.

Since TEs are potentially harmful for their host, transposition repression mechanisms have evolved in vertebrate genomes, which in turn favor elements that have the ability to escape repression. This process is exemplified in mammals where the *L1* element has repeatedly recruited novel 5' untranslated region to escape transcription repression by host factors [Adey et al., 1994; Khan et al., 2006; Sookdeo et al., 2013]. This molecular arms race was experimentally validated in primates, where *L1* is engaged in an arms race with the *ZNF91/93* transcriptional repressor [Jacobs et al., 2014]. In addition, the first open reading frame of *L1* evolves adaptively and is structurally unstable, suggesting that this protein may also be the site of an antagonistic interaction with a host factor [Boissinot and Furano, 2001; Khan et al., 2006; Sookdeo et al., 2013]. The evolution of *L1* in *A. carolinensis* does not fit this pattern. In fact the 5' untranslated region of the lizard *L1* is extremely conserved, even across highly divergent *L1* families, and the first open reading frame does not show any evidence of adaptive evolution or structural instability [Boissinot and Sookdeo, 2016]. This suggests that the control mechanisms of *L1* in *A. carolinensis* are not acting in a family-specific manner. We proposed that it is unlikely that family-specific transposition repressors evolved in an organism like the green anole, which harbors an extremely rich and diverse TE landscape, and that a more general repression mechanism constitutes a more effective way to control the large number of TE families that are simultaneously active in this genome [Boissinot and Sookdeo, 2016]. This aspect of reptile biology is however in dire need of investigation.

Since reptile genomes, and in particular squamate genomes, contain a large diversity of TEs (see below), they also contain a larger diversity of sequence motifs TEs can exchange. In mammals, *L1* elements have been shown to exchange genetic information by recombination, but this process occurs only between the same types of elements [Hayward et al., 1997; Sookdeo et al., 2013]. In contrast, the diverse repetitive landscape of reptiles offers many opportunities for exchange of sequence among TEs and for the generation of novel mosaic TEs. This is exemplified in the evolution of DNA transposons in green anole [Novick et al., 2011]. Some *hAT* elements contain fragments of other categories of TE, including SINEs and retrotransposons, and related *hAT* elements have exchanged some motifs repeatedly, probably by recombination. In an extreme case, some autonomous elements can carry up to 5 fragments derived from other TEs, accounting for as much as a third of the length of these elements. Thus, the evolution of TEs in anoles looks like a positive feed-

back loop in which the high diversity of TEs in this genome generates an even greater diversity of mosaic TEs that are exchanging fragments of sequence among themselves.

## The Diversity of Reptilian Mobilome

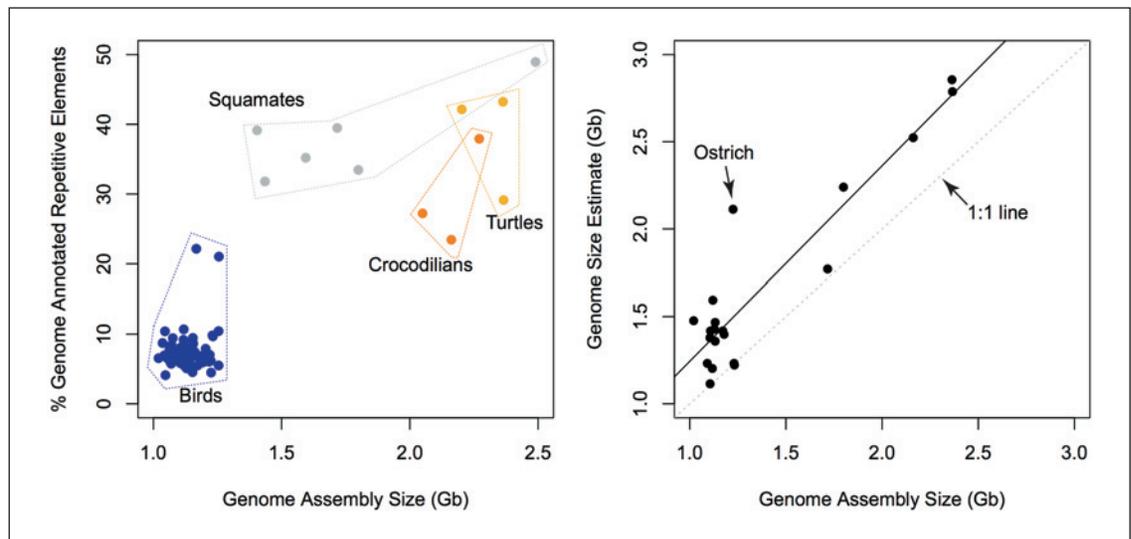
### *The Mobilome of Squamates*

Squamate reptiles are one of the most species-rich orders of vertebrates, with over 10,000 species ([www.reptile-database.org](http://www.reptile-database.org)), comprising all lizards and snakes, which together encompass over 200 million years of terrestrial evolution [Jones et al., 2013] and exhibit an incredible diversity of form, physiology, ecology, and behavior. Squamate reptiles provide many classic models of extreme adaptation, e.g., the remarkable growth and metabolic shifts observed in pythons [Andrew et al., 2015], multiple systems of venom evolution [Reyes-Velasco et al., 2015], repeated parallel instances of limb [Brandley et al., 2008] and sensory organ loss, and the remarkable adaptive radiations and convergent evolution documented in anole lizards, one of the most studied model systems in evolutionary biology [Losos, 2009]. Despite these fascinating aspects of squamate biology and their obvious utility as a research tool, only a limited number of high-quality squamate genome assemblies are available and have been analyzed for their repeat content. High-quality genome assemblies exist for the green anole *A. carolinensis* [Alföldi et al., 2011]; the Burmese python, *Python molurus bivittatus* [Castoe et al., 2013]; the king cobra, *Ophiophagus hannah* [Vonk et al., 2013], and the five-pace viper, *Deinagkistrodon acutus* [Yin et al., 2016] in addition to high-coverage assemblies of the Australian dragon lizard, *Pogona vitticeps* [Georges et al., 2015] and the Chinese crocodile lizard, *Shinisaurus crocodilurus* [Gao et al., 2017] as well as draft assemblies of the Burmese glass lizard, *Ophisaurus gracilis*, a legless anguid lizard [Song et al., 2015]; the leopard gecko, *Eublepharis macularius* [Xiong et al., 2016], and the corn snake, *Pantherophis guttatus* [Ullate-Agote et al., 2014]. This collection of genomes, when considered together, reveals that squamates are as diverse in TE content as they are diverse morphologically, physiologically, or ecologically. First, there is extensive variation in their TE content, ranging from 31.82% readily identifiable TEs (in the Burmese Python) to 49.6% (in the legless anguid lizard). While there have been extensive gene family expansions and contractions in squamate lineages, no clear correlations have been found linking these to TE content fluctuations [Yin et al.,

2016]. The diversity in genomic TE landscapes is best exemplified in snakes. A detailed comparison of the copperhead and Burmese python genomes showed that these 2 species differ considerably in their repeat content and the repeat age, although these 2 genomes are of relatively similar size [Castoe et al., 2011]. An analysis that included data from 10 snake genomes showed that TE families are at similar relative proportions in these genomes even though the total TE content varied substantially [Castoe et al., 2013]. Taken together, these results suggest that common mechanisms could be constraining and shaping squamate genomes in ways that we do not currently understand [Pasquesi et al., 2018] and that will likely only be elucidated by analyzing the evolutionary dynamics of TE families at the intraspecific level.

The most significant feature of squamate genomes is that they have a high diversity of TE families [Pasquesi et al., 2018] and, in some species, a large fraction of young and recently active elements. This TE diversity is most pronounced in the green anole genome, which has the most diverse representation of TE families observed in amniotes [Alföldi et al., 2011; Tollis and Boissinot, 2011; Chalopin et al., 2015]. In this species, all major groups of TEs are represented and many are active, including multiple non-LTR retrotransposons, LTR retrotransposons, endogenous retroviruses, DNA transposons, and *heli-trans*. A large proportion of identified elements are young, with little divergence from progenitor sequences, and a relative paucity of older degenerate sequences that predominate in mammalian and bird genomes. These qualities make the green anole genome more similar to that of fish than mammals [Novick et al., 2009] and support the hypothesis that the dynamic and diverse TE landscapes seen in squamate reptile genomes is likely more representative of the amniotes' basal state and that the limited TE activity seen in mammals and birds was separately and independently derived in each of these lineages.

Snake genomes appear to be intermediate in regards to genome diversity and activity. All major families (LINEs, SINEs, LTRs, and DNA transposons) contribute to snake genome size; however, in all cases, LINEs predominate and LTRs make the smallest contribution. Snake genomes contain a more mixed pallet of younger and older sequences than seen in the green anole genome. Older (divergent) copies of *L1* and *L2* occur in both basal and advanced groups of snakes, suggesting these elements were more active in ancestral snakes [Yin et al., 2016]. While recent expansions include the *snake1* CR1 LINEs that have primarily expanded in colubroid snakes [Castoe et al., 2013], expansion of DNA transposons (*hAT-Charlie*,



**Fig. 1.** Relationships of assembly genome size, genomic TE content, and genome size estimates. Left: The relationship of sequenced genome assembly size and TE genomic proportions for reptiles. Right: Discrepancies between genome size estimates from flow cell cytometry [Gregory et al., 2007] and sequenced genome assemblies in reptiles, for species in which both are available.

*Tc1/Mariner*) and LTR (*Gypsy*) sequences have occurred in the viper genomes and *L2* and *CR1* in boas and pythons [Yin et al., 2016]. Such recent TE expansions of elements suggest that the repetitive fraction of squamate genomes remains highly dynamic. This is further illustrated by the multiple horizontal transfer events that have been described from squamate genomes, including the independent horizontal transfer of *SPIN* elements within squamate genomes on more than 12 occasions over the past 50 million years [Gilbert et al., 2012]. Given that such events are primarily identifiable because the elements involved are also found in mammals, it is intriguing to imagine how many more such events might be identified with a greater diversity of genomes available.

#### *The Mobilome of Archosaurs*

The archosaur clade constitutes a major radiation within reptiles, including extant clades of birds and crocodiles as well as major lineages of extinct taxa including dinosaurs, distinct crocodylians, and pterosaurs [Nesbitt, 2011]. The basal split in archosaurs separates crocodylians (extant and extinct) from the clade including birds, dinosaurs, and pterosaurs. Birds are the most speciose group of archosaurs, with between 9,000 and 20,000 species depending on the species concept used for taxonomy [Barrowclough et al., 2016]. In contrast, nonavian dinosaurs may have had around 2,000 species at any given period of

time [Starrfelt and Liow, 2016], and crocodylians currently consist of only 20–30 species [Oaks, 2011].

The mobilomes of archosaur genomes vary to nearly the extent of the entire reptile clade (Fig. 1). Crocodylians exhibit a relatively high genomic TE prevalence, suggesting their genomes are more similar to the ancestral genomic TE landscape of reptiles [Wan et al., 2013; Green et al., 2014; Rice et al., 2017]. In contrast, at some point during evolution of the dinosaur lineage, genomic TE content became less prevalent [Organ et al., 2007]. This genomic reduction in TE content likely occurred in the lineage leading to the Saurischian dinosaurs, which includes extant birds and excludes pterosaurs and ornithischian dinosaurs. The genomic reduction of TEs in the Saurischian dinosaur clade is exemplified in the only extant taxon – birds – relatively lacking genomic TE diversity and abundance [Zhang et al., 2014; Chalopin et al., 2015; Gao B et al., 2017; Kapusta and Suh, 2017; Sotero-Caio et al., 2017] (Fig. 1).

Despite the large differences in total genomic TE content between crocodylians and birds, the retrotransposon superfamily *CR1* makes up the largest proportion of genomic TE content in both clades [Green et al., 2014; Zhang et al., 2014]. *CR1* elements comprise approximately 2–7% of most bird genomes [Zhang et al., 2014] and approximately 10% of crocodylian genomes [Green et al., 2014]. Additionally, crocodylian genomes are comprised

of large proportions of DNA transposons including *hAT* and *PIF-Harbinger* elements (~7% of the genome each) as well as LTR *Gypsy* elements (~3%) [Green et al., 2014]. While we cannot directly infer the genomic TE content and diversity of extinct archosaurs, we might expect a large proportion of *CR1* retrotransposons existed in dinosaurs and pterosaurs, given their prominence in extant archosaurs and activity and prevalence in reptiles in general.

While birds have a modest prevalence of genomic *CR1* content, their overall TE content is an outlier relative to other reptiles (Fig. 1). Birds have relatively small and conserved genome sizes relative to other reptiles [Gregory et al., 2007]. Small genome sizes in birds relative to closely related evolutionary lineages are likely due to metabolic constraints of powered flight [Gregory, 2001, 2002; Zhang and Edwards, 2012; Wright et al., 2014] and appear to be a convergently evolved trait among flying organisms, also including bats and pterosaurs [Organ and Shedlock, 2009]. The paucity of TE genomic activity and amplification in birds is a likely consequence of genome size constraint in birds. Recent work has hypothesized that TE activity in birds is accompanied by mid-scale (10s of kbp) deletions, with the overall result of largely stable genome sizes in birds despite TE turnover [Kapusta et al., 2017].

Crocodylians exhibit a relative paucity of TE activity similar to that seen in birds. The majority of *CR1* element activity predated the diversification of modern crocodylians [Suh et al., 2014]. However, in-depth analyses of crocodylian and bird genomes have uncovered interesting patterns of recent, but low, levels of activity. In crocodylians, there has been recent activity of endogenous retroviruses [Chong et al., 2014] and a novel class of SINEs [Kojima, 2015]. Similarly, recent activity and emergence of several TEs have been identified in multiple bird clades. Novel classes of SINEs are found in several avian lineages [Kapusta and Suh, 2017], and recent expansions of LTR and *CR1* retrotransposons have been identified in songbirds and woodpeckers, respectively [Zhang et al., 2014; Manthey et al., 2018; Suh et al., 2018]. These final 2 examples are notable: the downy woodpecker has about double the genomic TE content (>20% of the genome) relative to other birds [Zhang et al., 2014], and several flycatcher species (genus *Ficedula*) have thousands of LTR retrotransposon polymorphisms [Suh et al., 2018], indicating a moderate level of recent TE activity.

As more genomes are sequenced, and our ability to sequence, assemble, and analyze genomes improves, we will undoubtedly find we are currently underestimating the genomic TE content of archosaurs. While these genomes

will likely still have less TE content and activity relative to other reptiles, much of their genomes still need to be further explored and annotated. The future of genomic TE discovery in archosaurs is exemplified in 2 recent in-depth studies with manual repetitive element identification and better genome assemblies. Guizard et al. [2016] recently showed that the chicken genome was at least 19% simple sequence repeats and TEs, which nearly double its previous estimates of genomic repeat content. Similarly, better annotation and genome assemblies in a crow (genus *Corvus*) genome [Weissensteiner et al., 2017] identified large (>10 kb) satellite arrays that were missing from previous Illumina-based genome assemblies.

### *The Mobilome of Turtles*

Interest in the mobilome of turtles has emerged 30 years ago [Endoh and Okada, 1986] and has led to major contributions, such as how LINES and SINEs share their retropositional machinery by exchanging fragments of their sequences [Kajikawa et al., 1997]. Despite these early discoveries, little is known about the dynamics of TEs in this clade. This is likely due to the recent availability of genomic resources for turtles and the even more recent availability of extensive annotation. The first studies of turtle genomes concluded that 10% of the genome could be assigned to TE repeats [Shaffer et al., 2013; Wang et al., 2013]. This rather low proportion would place turtles as outliers within reptiles (Fig. 1), given the observed association between TE content and genome size in vertebrates [Chalopin et al., 2015; Elliott and Gregory, 2015]. These first reports however have likely underestimated the TE content in turtles' genomes, possibly due to a lack of de novo annotation of element families specific to reptiles. More recent annotations suggest that the 4 genomes currently sequenced – the green sea turtle (*Chelonia mydas*), western painted turtle (*Chrysemys picta belli*), Chinese soft-shell turtle (*Pelodiscus sinensis*), and Agassiz's desert tortoise (*Gopherus agassizii*) – display TE contents around 30% [Tollis et al., 2017]. This latter estimate would be more consistent with the general correlation between TE content and genome size observed in reptiles (Fig. 1).

LINES and DNA transposons dominate turtle genomes, and display a relatively uniform age distribution, suggesting slow but steady activity [Chalopin et al., 2015]. Among LINES, *CR1/L3* elements are the most abundant in turtles [Chalopin et al., 2015; Tollis et al., 2017], which is similar to other reptiles. The high number of *CR1* lineages in turtles likely reflects much of the *CR1* ancestral diversity in amniotes [Suh et al., 2014]. Many *CR1* subfamilies display a low divergence (<5%) from their con-

sensus sequence, suggesting a recent burst of activity [Tollis et al., 2017]. Overall, turtle genomes tend to diverge slowly [Shaffer et al., 2013] and do not display strong variation in size. Because conducting functional assays in these slow growing, slow to reproduce, and often endangered species seems especially prohibitive, extensive sequencing of this small group may be of particular utility for gaining a mechanistic perspective on the TE dynamics of turtles.

### The Impact of Transposable Elements on the Genome of Their Host

TEs are natural components of genetic diversity, and their activity and accumulation have profoundly impacted the evolution of genome size, structure, and function in vertebrates. Host genomes in turn have evolved mechanisms for defense against rampant TE amplification, including novel regulatory proteins, epigenetic silencing, and recombination-driven deletion [Nam and Ellegren, 2012; Goodier, 2016; Kapusta and Suh, 2017; Kapusta et al., 2017]. Host genomes have evolved mechanisms to control TEs because transposition is by definition mutagenic. Like any other mutation, TE insertions may have negative, neutral and nearly neutral, or positive consequences. These relationships between newly transposed TEs and host genomes are essentially symbioses, ranging from parasitism to commensalism and mutualism [Kidwell and Lisch, 2001]. When TE insertions are neutral, we expect genetic drift to be the main factor affecting the inserted TE's fate, similar to any neutral genetic variation. In contrast, TE insertions with negative or positive effects on host genomes will be under selective pressures. Negative selection will act on TE insertions with minor adverse effects on the host, while at the extreme, such as during rampant TE activity in cancerous cells, widespread TE insertions can have detrimental effects on the host organism and provide a dead end for TE amplification [Beck et al., 2011].

The relative importance of drift and selection in determining the fate of TE insertions has been examined in only one reptile, *A. carolinensis*. Since TE families tend to exhibit low divergence and ancient copies are conspicuously rare by comparison with mammalian genomes, it was proposed that TEs in lizard might be subject to a high rate of turnover [Novick et al., 2009; Tollis and Boissinot, 2011]. This model, first developed to explain the pattern of TEs' frequency in *Drosophila* and fish [Charlesworth and Charlesworth, 1983; Montgomery et al., 1987; Fura-

no et al., 2004], posits that novel insertions are so deleterious that they rarely reach fixation, so that mostly recent, low-frequency insertions should be found in populations. We tested this hypothesis by assessing LINE frequency in natural populations of green anole [Tollis and Boissinot, 2013; Ruggiero et al., 2017]. We found that, contrary to expectation, LINES do indeed reach fixation in anoles but that only short, truncated elements do so. Full-length copies, the only ones capable of further transposition, almost never reach fixation, and for those the turnover model does apply. Thus, LINE transposition does impose a substantial fitness cost in anoles, and in fact, the intensity of selection against LINES in anole may be stronger than it is in other organisms such as human and stickleback [Xue et al., 2018]. We also found that the fraction of elements that reached high frequency or fixation differed among populations [Tollis and Boissinot, 2013; Ruggiero et al., 2017]. For example, populations from North Carolina had significantly more fixed insertions in their genome than the population from Florida. This pattern can be explained by the demographic history of those populations [Tollis et al., 2012; Tollis and Boissinot, 2014; Manthey et al., 2016]. Northern populations of anoles have smaller effective population size than Floridian ones, and in those populations, the effect of drift was stronger and mitigated the effect of purifying selection, which was the dominant force in the south.

Although the majority of TE insertions will be either neutral or deleterious, TEs can occasionally be recruited by the host for its own benefit and thus positively selected [Warren et al., 2015]. There are now numerous examples of exaptation in vertebrates, and TEs have been shown to be a source of novel regulatory sequence (promoters or enhancers), of coding sequences (new genes or exons), of binding sites, and of noncoding RNAs (miRNAs and lncRNAs) [reviewed in Warren et al., 2015]. In fact, the exaptation of TEs has had some profound impacts on the evolution of networks as well as on the evolution of some important functions, such as immunity and neurotransmission [Feschotte, 2008; Jangam et al., 2017; Pastuzyn et al., 2018]. However, the vast majority of examples of exaptation in vertebrates have been found in mammals [Warren et al., 2015], arguably the best-studied vertebrate group in the field of comparative genomics. There is no reason to believe that the close examination of reptile genomes would not yield a trove of domesticated TEs. Indeed, the accumulation of TEs in the *Hox* clusters of reptiles [Di-Poi et al., 2009, 2010], which contrast with the extreme structural stability of those clusters in other vertebrates [Deschamps and Duboule, 2017], suggests that

TE insertions could have had an impact on the morphological diversity of squamates [Di-Poï et al., 2010] as well as the rate of speciation in anoles [Feiner, 2016]. Much work remains to be done to confirm a functional role of those insertions in the evolution of phenotypes.

While the above examples are all the result of transposition-induced mutations, post-insertion TE presence also has the potential to impact the structure and size of host genomes. When many TEs of similar sequence identity are spread throughout the genome, they may induce ectopic recombination resulting in insertions, deletions, and chromosomal rearrangements such as inversions [Canapa et al., 2015]. Thus, not only does transposition provide a source of structural diversity via TE insertions, but the TE insertions themselves may help create additional heritable structural variation affecting evolutionary processes across populations [Piacentini et al., 2014]. For example, TE-induced chromosomal changes may result in novel phenotypes [Saenko et al., 2015] or reproductive incompatibilities between isolated populations (e.g., Dobzhansky-Muller incompatibilities), lending a putative role for TEs in the speciation process [Brown and O'Neill, 2010; Rebollo et al., 2010].

A combination of TE-induced chromosomal changes as well as a heightened level of TE transposition would have profound consequences for populations and species. This situation could be induced via hypomethylation associated with cancers [Daskalos et al., 2009] or due to chromosomal rearrangements interrupting normal patterns of methylation and subsequently removing TE silencing [O'Neill et al., 1998; Dobigny et al., 2004]. Here, a cascade of genomic changes could promote evolutionary diversification and is consistent with several studies identifying coincident bursts of diversification and high TE activity and accumulation [Pascale et al., 1990; de Boer et al., 2007; Ray et al., 2008]. While genomic changes induced by TEs may promote speciation, the overall accumulation of TEs in the genome may have the opposite effect; lineages with larger genome sizes appear to have slower speciation rates relative to lineages with smaller genome sizes [Kraaijeveld, 2010].

As TEs accumulate in a genome, they naturally have the potential to increase genome size if transposition outpaces excision or deletion [Petrov, 2001]. TEs can increase genome size through copy number increase via transposition, TE-induced ectopic recombination resulting in changes in genome size, and by TE-mediated satellite expansions [Canapa et al., 2015; Klein and O'Neill, 2018]. While genomes may increase in size with TE activity, recombination-driven deletions counteract this ef-

fect, in some cases maintaining genome size even in the face of TE activity [Kapusta et al., 2017]. Although there is no obvious relationship between the complexity of organisms and the size of their genome, the accumulation of TEs can have significant structuring effects, which may be adaptive, for instance by promoting the establishment and maintenance of heterochromatin domains [Lyon, 1998; Slotkin and Martienssen, 2007; Kejnovsky et al., 2015]. In addition, differences in genome size can indirectly impact the adaptability of organisms by affecting the location and number of mutations of functional significance, e.g., “the functional space hypothesis” [Mei et al., 2018].

One might expect a simple relationship between genome size and genomic TE content. However, TE content varies across lineages and within genomes. TE content varies across the genomic landscape, especially in cases where species have microchromosomes such as in squamates and birds; microchromosomes are typically enriched in gene content and contain relatively fewer TEs [Ellegren, 2010]. Similarly, across the tree of life, clade-specific factors lead to the lack of a clear association between genome size and genomic TE content, while a strong association can be found within individual clades [Pagel and Johnstone, 1992; Canapa et al., 2015]. Based on published genomes, the reptile clade has a positive relationship between genome size and overall genomic TE content, including additional clade-specific patterns (Fig. 1). Overall, this pattern may be obfuscated by incomplete genome sequencing projects, where assemblies are underrepresenting the true size of genomes (Fig. 1).

## Conclusions

The sequencing of several genomes representative of the major reptilian lineages has considerably improved our understanding of amniotes genomic evolution and, in particular, revealed an extraordinary diversity of repetitive landscapes across reptiles. However, this field is only emerging, and most studies performed so far have been somewhat descriptive. There are still big gaps in our understanding: To what extent do TEs contribute to adaptation and speciation? Which mechanisms of repression have evolved in reptiles? How has this diverse repetitive landscape affected other genomic features such as the rates of recombination and mutation? How do different types of TE interact in their genomic environment?

To answer these questions, genome sequences of better quality are needed. In particular, chromosome-scale

genome assemblies are necessary. Poor assemblies often omit highly repetitive regions where TEs are more likely to accumulate. Without proper assembly and annotation, it becomes impossible to perform an exhaustive assessment of TE insertions. It can be assumed that many of the unassembled regions of genomes are largely, if not entirely, composed of repetitive sequences. The availability of chromosome-scale assemblies will help resolve this and possibly many aspects of TE dynamics. This will improve our estimates of TE content and may reveal impor-

tant trends in how mobile elements aggregate, are deleted, are domesticated, spawn simple repeats, and recombine into novel compound sequences. For similar reasons, further studies of TEs at the population level may be particularly useful. The breadth of reptile genomes and their TE content provides an ideal system for understanding the evolutionary forces that affect structural variation in natural populations and consequently identify general mechanisms that have shaped vertebrate genomic diversity.

## References

- Adey NB, Schichman SA, Graham DK, Peterson SN, Edgell MH, Hutchison CA 3rd: Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Mol Biol Evol* 11:778–789 (1994).
- Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, et al: The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477:587–591 (2011).
- Andrew AL, Card DC, Ruggiero RP, Schield DR, Adams RH, et al: Rapid changes in gene expression direct rapid shifts in intestinal form and function in the Burmese python after feeding. *Physiol Genomics* 47:147–157 (2015).
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, et al: Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–1310 (2002).
- Barrowclough GF, Cracraft J, Klicka J, Zink RM: How many kinds of birds are there and why does it matter? *PLoS One* 11:e0166307 (2016).
- Bartolomé C, Bello X, Maside X: Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol* 10:R22 (2009).
- Beck CR, Garcia-Perez JL, Badge RM, Moran JV: LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet* 12:187–215 (2011).
- Boissinot S, Furano AV: Adaptive evolution in LINE-1 retrotransposons. *Mol Biol Evol* 18:2186–2194 (2001).
- Boissinot S, Sookdeo A: The evolution of LINE-1 in vertebrates. *Genome Biol Evol* 8:3485–3507 (2016).
- Boissinot S, Chevret P, Furano AV: L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol Biol Evol* 17:915–928 (2000).
- Boissinot S, Davis J, Entezam A, Petrov D, Furano AV: Fitness cost of LINE-1 (L1) activity in humans. *Proc Natl Acad Sci USA* 103:9590–9594 (2006).
- Brandley MC, Huelsenbeck JP, Wiens JJ: Rates and patterns in the evolution of snake-like body form in squamate reptiles: evidence for repeated re-evolution of lost digits and long-term persistence of intermediate body forms. *Evolution* 62:2042–2064 (2008).
- Britten RJ, Kohne DE: Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161:529–540 (1968).
- Brown JD, O’Neill RJ: Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu Rev Genomics Hum Genet* 11:291–316 (2010).
- Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E: Transposons, genome size, and evolutionary insights in animals. *Cytogenet Genome Res* 147:217–239 (2015).
- Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning AP, et al: Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol Evol* 3:641–653 (2011).
- Castoe TA, de Koning AP, Hall KT, Card DC, Schield DR, et al: The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci USA* 110:20645–20650 (2013).
- Chalopin D, Naville M, Plard F, Galiana D, Volff JN: Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol Evol* 7:567–580 (2015).
- Charlesworth B, Charlesworth D: The population dynamics of transposable elements. *Genet Res* 42:1–27 (1983).
- Chong AY, Kojima KK, Jurka J, Ray DA, Smit AF, et al: Evolution and gene capture in ancient endogenous retroviruses – insights from the crocodylian genomes. *Retrovirology* 11:71 (2014).
- Cost GJ, Feng Q, Jacquier A, Boeke JD: Human L1 element target-primed reverse transcription in vitro. *EMBO J* 21:5899–5910 (2002).
- Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, et al: Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* 124:81–87 (2009).
- de Boer JG, Yazawa R, Davidson WS, Koop BF: Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8:422 (2007).
- Deschamps J, Duboule D: Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev* 31:1406–1416 (2017).
- Dewannieux M, Esnault C, Heidmann T: LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35:41–48 (2003).
- Di-Poi N, Montoya-Burgos JI, Duboule D: Atypical relaxation of structural constraints in *Hox* gene clusters of the green anole lizard. *Genome Res* 19:602–610 (2009).
- Di-Poi N, Montoya-Burgos JI, Miller H, Pourquié O, Milinkovitch MC, Duboule D: Changes in *Hox* genes’ structure and function during the evolution of the squamate body plan. *Nature* 464:99–103 (2010).
- Dobigny G, Ozouf-Costaz C, Waters PD, Bonillo C, Coutanceau JP, Volobouev V: LINE-1 amplification accompanies explosive genome re-patterning in rodents. *Chromosome Res* 12:787–793 (2004).
- Duvernell DD, Pryor SR, Adams SM: Teleost fish genomes contain a diverse array of L1 retrotransposon lineages that exhibit a low copy number and high rate of turnover. *J Mol Evol* 59:298–308 (2004).
- Ellegren H: Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol Evol* 25:283–91 (2010).
- Elliott TA, Gregory TR: What’s in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos Trans R Soc Lond B Biol Sci* 370:20140331 (2015).
- Endoh H, Okada N: Total DNA transcription in vitro: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. *Proc Natl Acad Sci USA* 83:251–255 (1986).

- Feiner N: Accumulation of transposable elements in *Hox* gene clusters during adaptive radiation of *Anolis* lizards. *Proc Biol Sci* 283: 20161555 (2016).
- Feschotte C: Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9:397–405 (2008).
- Feschotte C, Pritham EJ: DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41:331–368 (2007).
- Furano AV, Duvernell DD, Boissinot S: L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet* 20:9–14 (2004).
- Gao B, Wang S, Wang Y, Shen D, Xue S, et al: Low diversity, activity, and density of transposable elements in five avian genomes. *Funct Integr Genomics* 17:427–439 (2017).
- Gao J, Li Q, Wang Z, Zhou Y, Martelli P, et al: Sequencing, de novo assembling, and annotating the genome of the endangered Chinese crocodile lizard *Shinisaurus crocodilurus*. *Gigascience* 6:1–6 (2017).
- Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, et al: Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res* 17:992–1004 (2007).
- Georges A, Li Q, Lian J, O'Meally D, Deakin J, et al: High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *Gigascience* 4:45 (2015).
- Gilbert C, Schaack S, Pace JK 2nd, Brindley PJ, Feschotte C: A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464:1347–1350 (2010).
- Gilbert C, Hernandez SS, Flores-Benabib J, Smith EN, Feschotte C: Rampant horizontal transfer of SPIN transposons in squamate reptiles. *Mol Biol Evol* 29:503–515 (2012).
- Goodier JL: Restricting retrotransposons: a review. *Mob DNA* 7:16 (2016).
- Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, et al: Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346:1254449 (2014).
- Gregory TR: The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. *Blood Cells Mol Dis* 27:830–843 (2001).
- Gregory TR: A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class aves. *Evolution* 56:121–130 (2002).
- Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, et al: Eukaryotic genome size databases. *Nucleic Acids Res* 35:D332–D338 (2007).
- Guizard S, Piégu B, Arensburger P, Guillou F, Bigot Y: Deep landscape update of dispersed and tandem repeats in the genome model of the red jungle fowl, *Gallus gallus*, using a series of de novo investigating tools. *BMC Genomics* 17:659 (2016).
- Hartl DL, Lozovskaya ER, Lawrence JG: Nonautonomous transposable elements in prokaryotes and eukaryotes. *Genetica* 86:47–53 (1992).
- Hayward BE, Zavanelli M, Furano AV: Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*. *Genetics* 146:641–654 (1997).
- Ichiyanagi K, Okada N: Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. *Mol Biol Evol* 25:1148–1157 (2008).
- International Chicken Genome Sequencing Consortium: Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716 (2004).
- Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL: LINES between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. *Genome Biol Evol* 8:3301–3322 (2016).
- Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, et al: An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516:242–245 (2014).
- Jangam D, Feschotte C, Betrán E: Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet* 33:817–831 (2017).
- Jones ME, Anderson CL, Hipsley CA, Müller J, Evans SE, Schoch RR: Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuataras). *BMC Evol Biol* 13:208 (2013).
- Kajikawa M, Okada N: LINES mobilize SINES in the eel through a shared 3' sequence. *Cell* 111: 433–444 (2002).
- Kajikawa M, Ohshima K, Okada N: 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 (1997).
- Kapitonov VV, Tempel S, Jurka J: Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene* 448:207–213 (2009).
- Kapusta A, Suh A: Evolution of bird genomes—a transposon's-eye view. *Ann NY Acad Sci* 1389:164–185 (2017).
- Kapusta A, Suh A, Feschotte C: Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci USA* 114:E1460–E1469 (2017).
- Kejnovsky E, Tokan V, Lexa M: Transposable elements and G-quadruplexes. *Chromosome Res* 23:615–623 (2015).
- Khan H, Smit A, Boissinot S: Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res* 16:78–87 (2006).
- Kidwell MG: Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49–63 (2002).
- Kidwell MG, Lisch DR: Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* 55:1–24 (2001).
- Klein SJ, O'Neill RJ: Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res* 26:5–23 (2018).
- Kojima KK: A new class of SINEs with snRNA gene-derived heads. *Genome Biol Evol* 7: 1702–1712 (2015).
- Kordis D: Transposable elements in reptilian and avian (sauropsida) genomes. *Cytogenet Genome Res* 127:94–111 (2009).
- Kordis D, Gubensek F: Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc Natl Acad Sci USA* 95:10704–10709 (1998).
- Kordis D, Lovsin N, Gubensek F: Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. *Syst Biol* 55:886–901 (2006).
- Kraaijeveld K: Genome size and species diversification. *Evol Biol* 37:227–233 (2010).
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al: Initial sequencing and analysis of the human genome. *Nature* 409:860–921 (2001).
- Losos JB: *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* (University of California Press, Berkeley 2009).
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH: 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 (1993).
- Lyon MF: X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet* 80:133–137 (1998).
- Malik HS, Burke WD, Eickbush TH: The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* 16:793–805 (1999).
- Manthey JD, Tollis M, Lemmon AR, Moriarty Lemmon E, Boissinot S: Diversification in wild populations of the model organism *Anolis carolinensis*: a genome-wide phylogeographic investigation. *Ecol Evol* 6:8115–8125 (2016).
- Manthey JD, Moyle RG, Boissinot S: Multiple and independent phases of transposable element amplification in the genomes of Piciformes (woodpeckers and allies). *Genome Biol Evol* 10:1445–1456 (2018).
- Mei W, Stetter MG, Gates DJ, Stitzer MC, Ross-Ibarra J: Adaptation in plant genomes: bigger is different. *Am J Bot* 105:16–19 (2018).
- Montgomery E, Charlesworth B, Langley CH: A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet Res* 49:31–41 (1987).
- Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, et al: Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562 (2002).
- Nam K, Ellegren H: Recombination drives vertebrate genome contraction. *PLoS Genet* 8:e1002680 (2012).
- Nesbitt SJ: The early evolution of archosaurs: relationships and the origin of major clades. *Bull Am Mus Nat Hist* 352:1–292 (2011).
- Novick P, Smith J, Ray D, Boissinot S: Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. *Gene* 449: 85–94 (2010).

- Novick PA, Basta H, Floumanhaft M, McClure MA, Boissinot S: 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 (2009).
- Novick PA, Smith JD, Floumanhaft M, Ray DA, Boissinot S: The evolution and diversity of DNA transposons in the genome of the lizard *Anolis carolinensis*. *Genome Biol Evol* 3:1–14 (2011).
- Oaks JR: A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. *Evolution* 65:3285–3297 (2011).
- Ohshima K, Hamada M, Terai Y, Okada N: The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. *Mol Cell Biol* 16:3756–3764 (1996).
- O'Neill RJ, O'Neill MJ, Graves JA: Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393:68–72 (1998).
- Organ CL, Shedlock AM: Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol Lett* 5:47–50 (2009).
- Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV: Origin of avian genome size and structure in non-avian dinosaurs. *Nature* 446:180–184 (2007).
- Pace JK 2nd, Feschotte C: The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res* 17:422–432 (2007).
- Pace JK 2nd, Gilbert C, Clark MS, Feschotte C: Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci USA* 105:17023–17028 (2008).
- Pagel M, Johnstone RA: Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. *Proc Biol Sci* 249:119–124 (1992).
- Pascale E, Valle E, Furano AV: Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. *Proc Natl Acad Sci USA* 87:9481–9485 (1990).
- Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, et al: Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nature Commun* 9:2774 (2018).
- Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, et al: The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell* 172:275–288 e18 (2018).
- Pasyukova EG, Nuzhdin SV, Morozova TV, Mackay TF: Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. *J Hered* 95:284–290 (2004).
- Petrov DA: Evolution of genome size: new approaches to an old problem. *Trends Genet* 17:23–28 (2001).
- Piacentini L, Fanti L, Specchia V, Bozzetti MP, Berloco M, et al: Transposons, environmental changes, and heritable induced phenotypic variability. *Chromosoma* 123:345–354 (2014).
- Piskurek O, Okada N: Poxviruses as possible vectors for horizontal transfer of retrotransposons from reptiles to mammals. *Proc Natl Acad Sci USA* 104:12046–12051 (2007).
- Piskurek O, Nishihara H, Okada N: The evolution of two partner LINE/SINE families and a full-length chromodomain-containing Ty3/Gypsy LTR element in the first reptilian genome of *Anolis carolinensis*. *Gene* 441:111–118 (2009).
- Ray DA, Feschotte C, Pagan HJ, Smith JD, Pritham EJ, et al: Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res* 18:717–728 (2008).
- Rebollo R, Horard B, Hubert B, Vieira C: Jumping genes and epigenetics: towards new species. *Gene* 454:1–7 (2010).
- Reyes-Velasco J, Card DC, Andrew AL, Shaney KJ, Adams RH, et al: Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Mol Biol Evol* 32:173–183 (2015).
- Rice ES, Kohno S, John JS, Pham S, Howard J, et al: Improved genome assembly of American alligator genome reveals conserved architecture of estrogen signaling. *Genome Res* 27:686–696 (2017).
- Ruggiero RP, Bourgeois Y, Boissinot S: LINE insertion polymorphisms are abundant but at low frequencies across populations of *Anolis carolinensis*. *Front Genet* 8:44 (2017).
- Saenko SV, Lamichaney S, Martinez Barrio A, Rafati N, Andersson L, Milinkovitch MC: Amelanism in the corn snake is associated with the insertion of an LTR-retrotransposon in the *OCA2* gene. *Sci Rep* 5:17118 (2015).
- Schaack S, Gilbert C, Feschotte C: Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* 25:537–546 (2010).
- Shaffer HB, Minx P, Warren DE, Shedlock AM, Thomson RC, et al: The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol* 14:R28 (2013).
- Slotkin RK, Martienssen R: Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8:272–285 (2007).
- Smit AF, Tóth G, Riggs AD, Jurka J: Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* 246:401–417 (1995).
- Song B, Cheng S, Sun Y, Zhong X, Jin J, et al: A genome draft of the legless anguillid lizard, *Ophisaurus gracilis*. *Gigascience* 4:17 (2015).
- Sookdeo A, Hepp CM, McClure MA, Boissinot S: Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* 4:3 (2013).
- Sotero-Caio CG, Platt RN 2nd, Suh A, Ray DA: Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol Evol* 9:161–177 (2017).
- Starrfelt J, Liow LH: How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. *Philos Trans R Soc Lond B Biol Sci* 371:20150219 (2016).
- Suh A, Churakov G, Ramakodi MP, Platt RN 2nd, Jurka J, et al: Multiple lineages of ancient CR1 retrotransposons shaped the early genome evolution of amniotes. *Genome Biol Evol* 7:205–217 (2014).
- Suh A, Witt CC, Menger J, Sadanandan KR, Podsiadlowski L, et al: Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *Nat Commun* 7:11396 (2016).
- Suh A, Smeds L, Ellegren H: Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. *Mol Ecol* 27:99–111 (2018).
- Thomas J, Schaack S, Pritham EJ: Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol Evol* 2:656–664 (2010).
- Tollis M, Boissinot S: The transposable element profile of the anolis genome: how a lizard can provide insights into the evolution of vertebrate genome size and structure. *Mob Genet Elements* 1:107–111 (2011).
- Tollis M, Boissinot S: The evolutionary dynamics of transposable elements in eukaryote genomes. *Genome Dyn* 7:68–91 (2012).
- Tollis M, Boissinot S: 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 (2013).
- Tollis M, Boissinot S: Genetic variation in the green anole lizard (*Anolis carolinensis*) reveals island refugia and a fragmented Florida during the quaternary. *Genetica* 142:59–72 (2014).
- Tollis M, Ausubel G, Ghimire D, Boissinot S: Multi-locus phylogeographic and population genetic analysis of *Anolis carolinensis*: historical demography of a genomic model species. *PLoS One* 7:e38474 (2012).
- Tollis M, DeNardo DF, Cornelius JA, Dolby GA, Edwards T, et al: The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PLoS One* 12:e0177708 (2017).
- Ullate-Agote A, Milinkovitch MC, Tzika AC: The genome sequence of the corn snake (*Pantherophis guttatus*), a valuable resource for EvoDevo studies in squamates. *Int J Dev Biol* 58:881–888 (2014).
- Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, et al: The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci USA* 110:20651–20656 (2013).
- Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL: Widespread horizontal transfer of retrotransposons. *Proc Natl Acad Sci USA* 110:1012–1016 (2013).

- Wan QH, Pan SK, Hu L, Zhu Y, Xu PW, et al: Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res* 23:1091–1105 (2013).
- Wang Z, Pascual-Anaya J, Zadissa A, Li W, Nimura Y, et al: 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 (2013).
- Warren IA, Naville M, Chalopin D, Levin P, Berger CS, et al: Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosome Res* 23:505–531 (2015).
- Waters PD, Dobigny G, Waddell PJ, Robinson TJ: Evolutionary history of LINE-1 in the major clades of placental mammals. *PLoS One* 2:e158 (2007).
- Weissensteiner MH, Pang AWC, Bunikis I, Höijer I, Vinnere-Petterson O, et al: Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. *Genome Res* 27:697–708 (2017).
- Wright NA, Gregory TR, Witt CC: Metabolic “engines” of flight drive genome size reduction in birds. *Proc Biol Sci* 281:20132780 (2014).
- Xiong Z, Li F, Li Q, Zhou L, Gamble T, et al: Draft genome of the leopard gecko, *Eublepharis macularius*. *Gigascience* 5:47 (2016).
- Xue AT, Ruggiero RP, Hickerson MJ, Boissinot S: Differential effect of selection against LINE retrotransposons among vertebrates inferred from whole-genome data and demographic modeling. *Genome Biol Evol* 10:1265–1281 (2018).
- Yin W, Wang ZJ, Li QY, Lian JM, Zhou Y, et al: Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat Commun* 7:13107 (2016).
- Zhang G, Li B, Li C, Gilbert MT, Jarvis ED, et al: Comparative genomic data of the Avian Phylogenomics Project. *Gigascience* 3:26 (2014).
- Zhang Q, Edwards SV: The evolution of intron size in amniotes: a role for powered flight? *Genome Biol Evol* 4:1033–1043 (2012).
- Zupunski V, Gubensek F, Kordis D: Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol Biol Evol* 18:1849–1863 (2001).