The hidden elasticity of avian and mammalian genomes

Aurélie Kapusta[a], Alexander Suh[b] and Cédric Feschotte[a*]

[a] Department of Human Genetics, University of Utah School of Medicine, 15 N 2030 E, Salt - Lake City, Utah 84112, USA

[b] Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

*: corresponding author

ABSTRACT

Genome size in mammals and birds shows remarkably little interspecific variation compared to other taxa. Yet, genome sequencing has revealed that many mammal and bird lineages have experienced differential rates of transposable element (TE) accumulation, which would be predicted to cause substantial variation in genome size between species. Thus, we hypothesize that there has been co-variation between the amount of DNA gained by transposition and lost by deletion during mammal and avian evolution, resulting in genome size homeostasis. To test this model, we develop a computational pipeline to quantify the amount of DNA gained by TE expansion and lost by deletion over the last 100 million years (My) in the lineages of 10 species of eutherian mammals and 24 species of birds. The results reveal extensive variation in the amount of DNA gained via lineage-specific transposition, but that DNA loss counteracted this expansion to various extent across lineages. Our analysis of the rate and size spectrum of deletion events implies that DNA removal in both mammals and birds has proceeded mostly through large segmental deletions (>10 kb). These findings support a unified 'accordion' model of genome size evolution in eukaryotes whereby DNA loss counteracting TE expansion is a major determinant of genome size. Furthermore, we propose that extensive DNA loss, and not necessarily a dearth of TE activity, has been the primary force maintaining the greater genomic compaction of flying birds and bats relative to their flightless relatives.

INTRODUCTION

The nature and relative importance of the molecular mechanisms and evolutionary forces underlying genome size variation has been the subject of intense research and debate [1-5]. Variation in genome sizes may not always occur at a level where natural selection is strong enough to prevent genetic drift to determine their fate (neutral or effectively neutral variation) [3]. On the other hand, a number of correlative associations between genome size and phenotypic traits, such as cell size [6, 7] and metabolic rate associated with powered flight [8, 9], suggest that natural selection and adaptive processes also shape genome size evolution. Teasing apart the relative importance of these two forces (drift and selection) requires a better understanding of the mode and processes by which DNA is gained and lost over long evolutionary periods in different taxa. Thus, establishing an integrated view of the contribution of gain and loss of DNA to genome size variation (or lack thereof) remains an important goal in genome biology [e.g. 1, 2, 10, 11].

Most studies of genome size evolution have focused on taxa with extensive variation in genome sizes, such as flowering plants [12-17], conifers [18], insects [19-22], teleost fishes [23], or species with extreme sizes (such as pufferfishes [24, 25] and salamanders [26, 27]). Together, these investigations have documented that the differential expansion, accumulation, and removal of transposable element (TE) sequences represent a major determinant of genome size variation in plants and animals [see for reviews 2, 5, 28]. Generally, the studies cited above have revealed that species with larger genomes tend to have larger TE content combined with low rates of TE DNA removal, and *vice versa* for smaller genomes.

In comparison to these taxa, birds and mammals show little interspecific variation in genome size (from ~1 to 2.1 Gb and 1.6 to 6.3 Gb respectively, Figure 1), and little is known about the mechanisms underlying genome size homeostasis in these two classes of amniotes. In contrast to plants (and to a lesser extent fishes), changes in ploidy do not appear to represent a major source of genome size variation in birds or mammals, and there is no evidence of whole genome duplication events during amniote evolution [29]. On the other hand, it is well established that a considerable amount of new nuclear DNA has been generated throughout eutherian and avian evolution, mostly via TE expansion and, to a lesser extent, through segmental duplications [30-41]. These observations thus raise a

conundrum whereby TE activity has been pervasive in mammalian and avian evolution, yet has had apparently little impact on genome size.

The simplest way to reconcile this conundrum is to postulate that the amount of lineage-specific DNA gained by transposition has been systematically equalized by the removal of DNA along those lineages, thereby accounting for genome size homeostasis in mammals and birds [as hypothesized in 11, 42]. However, this hypothesis remains largely untested and, overall, little is known about the mode and tempo of genomic DNA loss in amniotes. An earlier comparative genomic study of insertion/deletion (indel) rates across 13 vertebrate genomes [43] implicated that variation in DNA gains through TE expansion acted in concert with variations in deletion rates to modulate genome size during evolution, but the region analyzed was limited to a 12-Mb alignment and included only a single bird species. A more comprehensive analysis of DNA gain and loss in the lineages of human, mouse and dog [33] showed that the dog and human lineages experienced 2.5x less DNA loss than in the mouse lineage, but also 2.8x and 1.6x less DNA gain, a balance explaining the modest differences in genome size across these species. Little is known about genome size dynamics in birds. Statistical models of genome size evolution in the avian lineage have inferred that a drastic contraction event (by ~0.8-fold) occurred prior to the divergence of birds in a theropod ancestor [44]. Consistent with this idea, a recent comparative analysis of 48 bird genomes revealed that introns and intergenic regions are, on average, ~2x and ~1.4x smaller in birds than in mammals and non-avian reptiles, respectively [40] [see also 45, 46]. Furthermore, the ostrich lineage was found to have experienced, on average, larger genomic deletions than alligator and turtle lineages [40]. Despite these recent insights, our understanding of genome size dynamics across eutherian and avian evolution remains fragmentary. In particular, the mode and tempo of DNA loss throughout amniote evolution has not been examined systematically.

Leveraging the recent sequencing of dozens of mammalian [e.g 47] and avian [40] genomes, we characterize genome size evolution in mammals and birds through an integrated analysis of DNA gain and loss on a genome-wide scale. The results are consistent with an 'accordion' model of genome size homeostasis whereby DNA gains are balanced by DNA loss, primarily through large-size deletions.

RESULTS

Genome size evolution as the integration of gain and loss of DNA

We used available genome assemblies for 10 eutherian (placental) mammals and 24 avian species, and their respective TE annotation (Methods). We first estimated the amount of DNA gained via transposition events in each lineage since their last common ancestor. For mammals, we compiled data from the literature as well as our own analyses to divide TE families previously characterized in each species into lineage-specific and ancestral families (Methods and Table S1). Using this information, we applied the RepeatMasker software [48] to infer the amount of DNA occupied (and therefore gained) by lineage-specific TEs in each of the genome assemblies examined. We added the amount of lineage-specific DNA gained by segmental duplications, when documented in the literature (information limited to some mammals) [31, 34, 35, 37, 38]. Because the evolutionary history of bird TE families has not been characterized as extensively as in mammals, we inferred the age of each TE insertion based on its divergence to the cognate family consensus sequence using lineage-specific neutral substitution rates [40] (Methods). For birds, gains were estimated from the DNA amounts corresponding to insertions younger than 70 My, which corresponds to the onset of the Neoaves radiation [49].

We then computed the total amount of DNA lost in each lineage by subtracting the amount of ancestral genomic DNA of each species (assembly size minus gains) to the "projected" assembly size of their common ancestor. It is important to note that our analysis requires using assembly sizes, which is the genomic space where TEs have been annotated, rather than actual genome sizes (which are always slightly larger due to current limitation in assembling highly complex regions such as centromeres). For eutherians, we used an ancestral genome assembly size previously estimated at 2.8 Gb based on a multiple alignment of 18-species genome assemblies, allowing ancestral reconstruction [50, 51]. For birds, we used 1.3 Gb as the predicted assembly size for both the ancestor of Paleognathae and Neoaves based on ancestral genome sizes previously inferred for these two clades [52] and a comparison of genome sizes and assembly sizes for each of the bird species sequenced (Methods). Using these inferences, we estimated the total amount of DNA lost along each of the 34 lineages considered (Dataset S1).

As an example of the approach, in the human lineage we estimated that 899 Mb of the hg38 assembly consisted of DNA gained via lineage-specific TE insertions (815 Mb) and segmental duplications (84 Mb), which leaves an "ancestral DNA" amount in the human genome assembly of 2,150 Mb. Thus, we can infer that 650 Mb (2,800 minus 2,150) of ancestral DNA was lost in the human lineage over the past ~100 My. The same procedure was applied to the other species lineages considered. Based on these amounts, we computed lineage-specific DNA loss coefficients k [as in 33] with $E = A\,e^{-kt}$, where E is the amount of extant ancestral sequence in the species considered, A the total ancestral assembly size, and t time (100 My for mammals and 70 My for birds) (Methods and Dataset S1). Applying our predicted DNA loss rate for human, we obtain a coefficient k of 0.0026 per My, which is nearly identical to the previously calculated coefficient of 0.0024 [33], despite being based on a different methodology to infer DNA gain and loss.

When applied to all lineages (Figures 2, 3 and S1), the results of these analyses show that the amount of DNA gained and lost has varied substantially across lineages. DNA gains vary by more than 6-fold across mammals (from 150 Mb in the megabat to 1,007 Mb in the mouse lineages, Figure 2) and by more than 30-fold across birds (from 7 Mb in the ostrich to 255 Mb in the woodpecker lineages, Figure 3). DNA loss amounts range by 2-fold across mammals (from 650 Mb in the human to 1,373 Mb in the microbat lineages), and by more than 3-fold across birds (from 119 Mb in the ostrich to 424 Mb in the woodpecker lineages).

For mammals, these results confirm the trends previously reported for some of these lineages [30, 31, 33]: we observe more gains in rodents than in human, and more gain in human than in dog, together with more loss in rodents than in dog or human. In addition, we found that the coefficient at which DNA was lost along the lineages examined is the lowest in the ostrich lineage and the highest in the microbat lineage (Figures 2 and 3). In fact, both the microbat and megabat lineages stand out as having both the highest amount and coefficient of DNA loss, followed closely by the mouse and rat lineages (Figures 2 and 3). Altogether, we found that neither DNA gain nor loss can solely explain variation in genome (assembly) sizes among the mammals and birds examined (Dataset S1). These results imply that variations in DNA gain and loss have acted in concert to modulate genome size in eutherian and avian evolution. We note that loss exceeded gain in all but two lineages (human and elephant, Figures 2, 3 and S1).

To investigate the extent by which these two opposite forces each contribute to genome size homeostasis, we next examine whether DNA gains (percentage of the ancestor assembly size) and DNA loss coefficients correlate with assembly sizes. In mammals, we observe a significant positive correlation between DNA gains and assembly sizes, while DNA loss are negatively (but non-significantly) correlated with assembly sizes ($R^2 = 0.73$ with $p = 0.02$ and $R^2 = 0.49$ with $p = 0.15$, Dataset S1). However, these results have to be interpreted with caution because our statistical power is limited by the relatively small sample of mammal lineages examined (n=10). For the 24 bird lineages analyzed, we observe that DNA loss coefficients, but not DNA gains, significantly correlate with assembly sizes ($R^2 = 0.63$ with $p = 0.0005$ and $R^2 < 0.1$ respectively). This observation indicates that DNA loss is a major contributor to genome size homeostasis in birds. However, neither one of these two forces alone can fully account for the variation between extent assembly sizes (e.g woodpecker, Dataset S1). Thus, these two forces must have acted in concert to modulate genome size throughout avian evolution. Consistent with this idea, we found that DNA gains and DNA loss coefficients are positively and significantly correlated with each other across the bird lineages examined (Pearson coefficient $R^2 = 0.65$, $p = 0.0005$, Figure 4). These data support a model where genome size homeostasis is maintained through DNA loss counteracting DNA gains through TE expansion. This is most strikingly illustrated in the woodpecker lineage, which shows both the largest amounts of gains and the highest DNA loss coefficients (Figures 3 and S1, Dataset S1).

Contribution of microdeletions to DNA loss

Having determined that DNA loss makes an important contribution to eutherian and avian genome size homeostasis, we next sought to investigate the types of deletion events involved in the process. We first assessed the impact of small deletions (< 30 bp; hereafter 'microdeletions') through multispecies alignment available from the UCSC genome browser (MultiZ output of 100-species alignments). We separately extracted and analyzed genomic alignments for 11 eutherian species (plus the marsupial *Monodelphis domestica* as outgroup) and for seven bird species (plus the lizard *Anolis carolinensis* as outgroup). We used the principle of parsimony to infer and place microdeletion events (estimated as sequence gaps of less than 30 bp) on the phylogenetic tree of the species (Methods). The total amount of concatenated alignment analyzed in this way corresponded to 237 Mb and 52 Mb for mammals and birds, respectively. Microdeletion rates were obtained for each lineage by normalizing the amount of gaps in the alignment by its total length and dividing

the amount of alignment gaps per lineage by the corresponding branch length in My (Figure 5).

The results show that rodents have the highest microdeletion rates among the lineages examined (Figure 5A), about 3.5x higher as those for the human lineage, in agreement with previous analyses of a smaller number of mammals [26, 30, 31, 53, 54]. All other mammal lineages we analyzed exhibit microdeletion rates that are intermediate between those of human and rodents, except for the common ancestor of bats (Chiroptera) which displays the lowest microdeletion rate in our analysis. Overall, microdeletion rates do not appear to vary substantially within a given mammalian (super)order (Primates, Rodentia, Chiroptera, Carnivora). One exception is Afrotheria, where the elephant lineage is characterized by a much lower (0.4x) microdeletion rate than that of the tenrec (*E. telfari*) or that of their common ancestor. This could be linked with peculiar characteristics evolved in the elephant lineage, such as large body size, long gestation, slow development to maturity, long generation time [see for review 55]. When the number of microdeletion events (normalized for alignment length) is considered, rather than the total amount of DNA removed by microdeletions (Figure 5A), we observe similar trends whereby bat, primate, and rodent lineages display the lowest, second lowest, and highest microdeletion rates, respectively.

Among the seven bird species considered in the alignment, we found that the two finches display the highest microdeletion rates (both in amount of DNA removed and number of events), which are 5.4x higher as those in the falcon lineage. All other bird lineages show rates and number of events that are intermediate between those of finches and falcon. We find that the average length of microdeletions in birds (5.8 bp) is slightly larger than inferred in mammals (5.2 bp) (Dataset S2). Average lengths per species lineages range from 4.8 bp (bats) to 5.7 bp (tenrec) in mammals and from 5.5 bp (zebra finch) to 6.1 bp (rock pigeon) in birds, and the distribution profiles are significantly different in pairwise species comparisons between birds and mammals, except between the zebra finch and some of the mammals (Dataset S4 and Figure S2). Within orders, pairwise species comparisons are significantly different between bats and other mammals, and between zebra finch and the other birds (Dataset S4 and S2). Microdeletions are slightly smaller in size in human and macaque than in mouse and rat (by 0.46 bp on average), which is in agreement with a previous study [54]. In summary, we observe substantial variation in the rate and size spectrum of DNA microdeletion between and within eutherian and avian orders. This may reflect species

characteristics such as mechanisms at the origin of microdeletions and effective population size [56] (see Discussion).

Next, we applied the microdeletion rates inferred over the various branches of the phylogeny to extrapolate the amount of DNA lost through this class of deletion during the last 100 My in mammals and the last 70 My in birds. We compared this to the total amount of DNA estimated to be lost within the same timeframe (Figures 2, 3 and S1). The results indicate that microdeletions account for only a small fraction (<10%) of the DNA lost in mammals and birds over the past 100 My and 70 My respectively (from 1% for the chicken lineage to 8.2% in the rat lineage; Dataset S2).

Contribution of midsize deletions to DNA loss

The method above enables the extrapolation of microdeletion rates over the various branches of the tree, but relies on a multispecies alignment using human as a reference. Therefore, it inherently favors the retention of regions alignable between human and the species considered, potentially biasing our analysis for the most conserved regions of the genome. Additionally, large deletions are, by design, excluded from the MultiZ alignment blocks used in the analysis above. In order to estimate microdeletion rates in a less biased fashion, as well as to capture larger deletion events, we developed an independent approach relying on the comparison of closely related species. We designed our computational pipeline to capture deletions up to a specified length (10 kb) in trios of species representative of the primate, chiropteran (bats), carnivore, artyodactyl, and afrotherian lineages, as well as eight trios of bird species (Figure 6). Briefly, the approach selects at random a pair of anchor sequences separated by a set length in the genome of the outgroup species as a query in BLAT searches to identify orthologous regions in the other two species (Methods). Three-way species nucleotide sequences alignments are then generated and deletions are quantified from the amount of gaps in the alignment and parsimoniously placed along the branches of the phylogeny (Methods). Deletion rates were obtained by dividing the total length of alignment gaps per given species lineage by the corresponding branch length in My. To be able to compare across lineages, the deletion rates were also normalized by alignment length (shown in Mb in Figure 6).

We first used these datasets as an independent method to infer microdeletion (<30 bp) rates along the lineages represented and compared the results with those obtained with the

MultiZ approach outlined above. We observe that the trends observed using the MultiZ approach for this subset of species are largely recapitulated (Figure 6, light green circles). For instance, the elephant displays the lowest microdeletion rates among the species compared, while bats and medium ground finch display the highest. However, microdeletion rates inferred by this method are consistently higher than those estimated based on the MultiZ alignment, on average 1.6 times higher for mammals and 4.3x higher for birds (Dataset S3). Presumably this difference reflects the greater evolutionary constraint of the genomic regions aligned by MultiZ, which generally leads to an underestimation of the "neutral" microdeletion rates of the species. Nevertheless, the fraction of total DNA loss accounted for by microdeletions when applying the new rate estimates remain modest, ranging from 5.1% in the cow lineage to 15.4% in the medium ground finch lineage (Dataset S2). Thus, the vast majority of DNA loss during eutherian and avian evolution must have occurred through deletions larger than 30 bp.

We then sought to capture deletions larger than 30 bp, and an analysis of the size spectrum of alignment gaps recovered shows that our computational pipeline succeeded in capturing relatively large deletion events (Figure S3). For example, in the human lineage, 5.8% of the gaps were longer than 1 kb and the largest deletion event identified was 9,022 bp relative to the macaque genome (breakpoint in hg38 at chr10:81816824). Overall, between 10 and 23% of the deletion events recovered in each species were larger than 200 bp (Figure S3). We manually verified the longest events recovered in each species: in mammals the largest was a ~10-kb deletion in the chimpanzee relative to the macaque genome (breakpoint in panTro4 at chr14:92050068). In birds, the largest event corresponded to a ~6-kb deletion in the medium ground finch relative to the golden-collared manakin (see legend of Figure S3 for the breakpoint coordinates of all longest events). Because longer deletion events tend to be fragmented into several gaps when the sequences of the three species are aligned, counting the number of gaps in the alignment would likely overestimate the actual number of deletion events in each species lineage. Thus we focused our analysis of midsize (>30 bp) deletion events on the *amount* of DNA lost through this class of deletion, rather than the actual number of events.

The results across the five eutherian orders examined (Figure 6) reveal trends similar to our analysis of microdeletion rates, with the elephant and the microbat showing the lowest and highest rates of midsize deletions, respectively (Figure 6, dark grey circles). By applying the

rates of midsize deletion inferred for each mammalian lineage to the entire distance separating these species from their common ancestor (~100 My ago), we were able to estimate that the amounts of DNA lost via this class of deletion ranged from 62 Mb in the elephant lineage to 134 Mb in the human lineage. These extrapolated figures (Dataset S3) suggest that midsize deletions have accounted for 7.3% (elephant) to 20.7% (human) of the total amount of DNA lost during eutherian evolution (11% on average). Together, micro- and midsize deletions account for about 30.9% of DNA loss in the human lineage, and only 14.1% in the microbat lineage (13% in the elephant lineage, 18% on average, Dataset S3). These data suggest that the vast majority of eutherian DNA loss has occurred through deletion events larger than those we are able to capture here (~10 kb).

For birds, the results of our midsize deletion analysis reveal trends similar to those for microdeletions as well: the medium ground finch, Anna's hummingbird, and woodpecker lineages show the highest rates, while the smallest rates are observed in the ostrich and penguin lineages (Figure 6 and Dataset S3). Next, we sought to assess the relative contribution of micro- and midsize deletions in birds to genome size homeostasis. We took advantage of the statistical power enabled by the analysis of 16 species to test the relationship between the rates of these two classes of deletions and the DNA loss coefficients calculated over 70 My of evolution (Dataset S3). We observe that the two variables are positively correlated, either if each class of deletion is considered separately ($R^2 = 0.79$ with $p = 0.0003$ and $R^2 = 0.63$ with $p = 0.009$ for micro- and midsize deletions respectively, Dataset S3) or together ($R^2 = 0.73$ with $p = 0.0015$, Dataset S3). These results suggest that both classes of deletions are significant drivers of genome size homeostasis during avian evolution.

The extrapolation of the amount of DNA lost in birds through the combined action of micro- and midsize deletions vary by up to one order of magnitude across lineages, from 14 Mb (ostrich and Adelie's penguin) to 136 Mb (woodpecker) (Dataset S3). These amounts account from 9.5% (Adelie's penguin) to 39.2% (Anna's hummingbird) of the total DNA loss along the bird lineages examined (21.7% on average). Thus, on average, the contribution of two classes of deletions is similar in birds and mammals, but we observe a greater variation in micro- and midsize deletion rates among birds than among mammals.

DISCUSSION

DNA gain and loss analysis reveals the elasticity of avian and mammalian genomes

Our study represents, to our knowledge, the most systematic analysis of the amount of genomic DNA gained and lost during eutherian and avian evolution, two taxa showing remarkably little interspecific variation in genome sizes compared to others (Figure 1). One interpretation for this apparent stasis in genome size could be that these lineages simply experienced relatively small amounts of DNA gain and loss during evolution [57-59]. Our analysis shows that this is clearly not the case: there has been extensive DNA gain and loss of DNA throughout eutherian and avian evolution. For example, the amount of DNA gained via lineage-specific transposition in the mouse lineage contributed to a net gain of DNA equivalent to 33% of the current genome content, while the equivalent of 44% of genome content was lost over the same timeframe (Figure 2 and Dataset S1). The woodpecker lineage provides another striking example. Amongst birds, this species lineage has experienced the largest amount of DNA gain (255 Mb, predominantly through CR1 LINE transposition [40]) but also the largest amount of DNA loss (424 Mb, equivalent to about a third of the genome) over the past ~70 My, resulting in a current genome size comparable to that of other modern bird species (Figures 3 and S4, Dataset S1). Thus, our data reveal a previously underappreciated level of elasticity in eutherian and avian genomes.

These findings allow us to uncover a general pattern of genome evolution along the major avian and eutherian lineages, whereby the (often large) amount of DNA gained via lineage-specific transposition is essentially balanced by the amount of DNA lost over the same timeframe. This 'accordion' process helps explaining the relative maintenance of genome size across the eutherian and avian phylogeny. This is particularly evident in birds (Figure 1), which display a positive correlation between DNA gain and DNA loss (Figure 4). Thus, our results indicate that the relatively small genome size of birds is not merely due to a dearth of transposition in those lineages, as previously hypothesized, [e.g. 57, 58, 59], but rather the result of a dynamic interplay between TE-mediated DNA acquisition and subsequent DNA loss [as suggested in 11, 42].

DNA loss through large deletions as a determinant of genome size

Previous studies assessing DNA loss have mainly focused on deletions within TE sequences, which imposes a relatively small upper limit for the size of observable events

(since TE copies rarely exceed 10 kb). The rate of deletions estimated through this approach have been shown to be a major predictor of genome size evolution in insects [e.g. 19, 20-22, 60], plants [e.g. 13], and a few vertebrates [11, 24-27]. However, whether the variation in the rate of small deletions can actually account for genome size variation observed between taxa has been questioned [discussed in 61, 62]. Indeed, quantifications from limited comparative datasets have suggested that microdeletions alone cannot account for the extent of genome contraction observed in some vertebrate lineages [e.g. 2, 23, 53, 54, 63]. Here, we assessed a broader size spectrum of deletion through whole-genome and local alignments of diverse birds and mammals. Our estimates of microdeletion (1-30 bp) rates show that this type of events can only explain a minute fraction of the DNA content lost during avian and eutherian evolution (Figure 5 and Dataset S3) and as such do not appear to be a major contributor to genome size evolution in these taxa.

Our results show that midsize deletions (31 bp to 10 kb) play a larger role than microdeletions in explaining the observed interspecific variation in DNA loss. Collectively, however, micro and midsize deletions detected in our analyses still account for a limited fraction (9.5 to 40% and 20% in average, Dataset S3) of total DNA loss in eutherian and avian evolution. These data suggest that the vast majority of DNA loss in amniotes has been driven by relatively large deletions (>10 kb). Such large deletions are challenging to detect systematically with currently available genome assemblies, precluding us to measure the rate of these events along the lineages considered in this study. We note, however, that instances of large chromosomal deletions have been documented previously in mammals (e.g. 1,511 and 845 kb [64] and 31 kb [65]; see also [66]), and we were able to identify individual events (Figure S5). Similarly, large segmental deletions were inferred to have had occurred in the common ancestor of birds (118 events for a total of 58 Mb and up to 2.1 Mb per event [40]).

Such large deletions, combined with the sheer amount of DNA loss in some of the mammal and bird lineages examined (up to 37.9% and 22.6% of nuclear DNA content respectively), underscores the dispensability of a large fraction of genomic DNA in these animals [66-68]. Yet it does not preclude that the process of segmental DNA loss has played an important role in driving phenotypic evolution [69-71]. In fact, there are strong hints that large deletions caused a substantial level of gene loss in birds (274 protein-coding genes) with potentially profound phenotypic consequences [40, 72, 73]. The foreseeable improvement of genome

assembly via third-generation sequencing (e.g. long-read sequencing and gap filling) [see 74] will provide a way to more directly test the hypothesis that large deletion events play a prominent role in amniote genome evolution.

Genome contraction covaries with transposable element expansion

What could be the mechanisms facilitating the co-variation between the amounts of DNA gained via TE insertions and the DNA that is lost, which is especially striking in bird evolution (Figure 4)? One simple explanation would be that TE insertions and deletions occur and fix at comparable rates in a given species lineage because they are governed by the same population genetics parameters, which also govern variation in other, largely neutral mutational processes such as nucleotide substitutions [1, 56]. Consistent with this idea, we find that microdeletion rates (and, to a lesser extent, midsize deletion rates) correlate strongly with neutral substitution rates in birds (Datasets S2 and S3). This may suggest that variation in microdeletion rates largely reflects population genetic parameters, with large effective population sizes leading to an uncoupling of neutral genetic variation from nearby deleterious alleles, i.e., a less pronounced effect of linked selection [75]. Conversely, natural selection acts less efficiently in species with smaller effective population sizes [75], which has been suggested to contribute indirectly to the reduced purging of slightly deleterious TE insertions [3, 76, contra 77]. Mammals have generally smaller effective population sizes than birds [78], which is predicted to increase the probability of fixation of nearly-neutral TE insertion and deletion events through genetic drift [75, 79]. This prediction is congruent with our observation that the overall amounts of DNA gained and lost have been more substantial in mammals than in birds (Figures 2, 3 and S1).

Mechanistically, the circumstances of frequent fixation of TE insertions would also provide a plausible fodder for large chromosomal deletions. Indeed, interspersed repeats with high level of sequence similarity such as recently expanded TE families represent a prime substrate for non-allelic homologous recombination (NAHR) events that may result in the deletion of the intervening DNA (reviewed in [80]. While the impact of TE-mediated NAHR on the process of DNA loss has been well-documented in plants [e.g. 12, 15, 16, 81-84] [see for review 28], it has not been systematically examined in vertebrates. Nonetheless, comparative studies in primates have suggested that an increased density of TEs from the same family augments the probability of inter-element NAHR deletion events to occur between TE copies. For example, the highly abundant *Alu* elements have mediated

considerably more NAHR deletion events [e.g 65, 85, 86] than L1 [87] or SVA [88] elements, which occur at much lower density in primate genomes. Thus, it is tempting to speculate that the explosive amplification of one or a few TE families, such as CR1 elements in woodpeckers [40] and Ves SINEs in bats [89, 90], led to an increase opportunity for NAHR thereby facilitating the extreme degree of DNA loss that we observed in these two lineages (Figures 2, 3 and S1). The idea that genome expansion via transposition subsequently promotes genome contraction via large-scale TE-mediated deletion would provide a mechanistic underpinning for the proposed 'accordion' model of genome size evolution.

Implications for the origin of flight in amniotes

Overall, our findings provide support for a general trend of strong genome contraction throughout the evolution of bats and birds (Figures 2, 3, S1 and S4), the only vertebrates capable of powered flight, which extends the previous notion that the evolution of the small genomes of bats and birds predates the emergence of flight [44, 91]. The continuous genome contraction we see along multiple bird lineages (Figures 3, S1 and S4) is consistent with previous inference that their common ancestor had a larger genome than that of extant avian species [52]. Importantly, we found no significant elevation in microdeletion rates in the respective common ancestor of bats, Palaeognathae (ratites and tinamous), Galloanserae (chickens and ducks), or Neoaves (all remaining birds) (Figure 5). Paradoxically, bats display the lowest microdeletion rate in our analysis (Figure 5A). In birds, our results are in agreement with previous estimates of rates of deletions <100 bp in the ancestors of Aves, Neognathae and Neoaves (~0.3, 0.4 and 0.2 Mb per My, [Fig S12 of 40]). Together, these observations suggest that genome contraction prior to the evolution of flight in the common ancestor of birds and bats must have occurred through relatively large chromosomal deletion events, but not through an increased rate of microdeletions.

Genome size variation between bird species has been linked to variation in metabolic cost of powered flight, with hummingbirds exhibiting the highest metabolism and smallest genomes [9, 92, 93], while flightless ratite birds display the largest genomes [2, 45, 94]. Our results lend further support to this connection between metabolic rate and constraint in genome size. We found that bird lineages that have lost flight (penguins and ostrich) are characterized by midsize deletions rates significantly lower than those of flying birds (2.3-fold on average; ks test, p = 0.0036; Figure 6, Dataset S3). This trend is also consistent with the results of a recent study indicating that TE removal through ectopic recombination

occurs at a faster rate in the zebra finch (flying bird) than in the chicken (ground-dwelling bird) [42]. Furthermore, we observe that flightless bird lineages (penguins and ostrich) have gained generally less DNA during evolution than flying birds (Figures 3 and S1, Dataset S1). Thus, the larger genomes of flightless birds do not appear to reflect increased DNA gains, but slower removal of DNA relative to flying birds. In other words, the genomes of flightless birds are less dynamic overall than those of flying species.

In addition to their connection with powered flight, resting metabolic rates are correlated with body mass in mammals [95] and birds [96]. Interestingly, in our dataset we note that animals larger than other species within the same order (e.g. elephant vs. manatee and tenrec, cow vs. sheep, ostrich vs. tinamou) display lower micro- and midsize deletion rates (Figures 5 and 6). Similarly, megabats have larger body mass than microbats, and show a lower DNA loss coefficient (Figure 2). These observations are consistent with a relationship between body mass, resting metabolic rates, and genomic deletion rates. However, we did not detect any general correlation between body mass and DNA loss when all mammals and birds in our dataset are considered (Datasets S1 and S3), suggesting that there is no simple relationship between these parameters.

Finally, our results in bats are also consistent with the hypothesis that the metabolic requirements for powered flight constrain genome size [91, 97] (Figure 1). We found that bats have a DNA loss/gain ratio ~4.3-fold higher than the other mammals examined (Figure 2), as well as the highest midsize deletion rates (Figure 6, Dataset S3). Importantly, however, we observe that neither bats (Figures 5A and 6), nor flying birds (Figure 6 and Dataset S3) exhibit increased microdeletion rates relative to their flightless outgroups, again implying a predominant role of large deletion events in keeping the genomes of the flying species particularly streamlined. Further studies are warranted to better characterize the molecular mechanisms underlying these large chromosomal deletions and their biological significance in amniote evolution.

METHODS

Genomic and biological data

Version of assemblies and species names are listed in supplementary datasets. Genome sequences in fasta format were recovered from UCSC for mammals (http://hgdownload.soe.ucsc.edu/goldenPath) and from ftp://climb.genomics.cn/pub/10.5524/100001_101000/101000/assembly/ [40] for birds. Body mass data are from [95, 98, 99] for mammals, and from [100] for birds.

TE annotation

For all mammals besides bats, RepeatMasker outputs were downloaded from http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.htmlLibraries, RepeatMasker open-4.0.5 [48] ran with the repeat library release 20140131 from repbase [101]. For bats, we obtained the TE annotation with RepeatMasker open-4.0.5 (using -e ncbi), using a custom library based on TE consensus sequences available in the literature [89, 90, 102-106]. For birds, we used available RepeatModeler outputs of 44 bird species [107], ran RepeatModeler locally for the duck genome, manually curated selected repeats and complemented these libraries with all available avian TE annotations [32, 101, 108]. All sequences were merged in one unique library using a custom Perl script (ReannTE_MergeFasta.pl, available at https://github.com/4ureliek/ReannTE). We obtained TE annotations by running RepeatMasker open-4.0.5 on all genomes of interest with the same library (using -e ncbi).

Determination of ancient versus lineage specific TEs

For mammals, we classified TE families as lineage-specific or shared between placental mammals (Table S1). We compiled data from Repbase and annotations of the RepeatMasker libraries [48, 101], complemented by our own orthology assessment (combination of BLAT, observation of the conservation tracks on UCSC, and orthology assessment with the following script: https://github.com/4ureliek/TEorthology). In birds, the majority of TEs belong to the CR1 superfamily [32, 40]. CR1 have been active at least since the common ancestor of birds, with always several subfamilies at the same time [109]. This is because one CR1 lineage survived from the many lineages of LINE present in the common ancestor of birds and crocodilians [110]. CR1 consensus sequences tend to be similar between ancient and recent families (e.g., families CR1-E and CR1-J across most of

avian evolution [109], which creates mis-annotations in the genome using RepeatMasker. Therefore, we relied on substitution rates to split TE-derived DNA into lineage-specific or shared. We developed a Perl script (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs/blob/master/parseRM.pl) to parse the raw alignment outputs from RepeatMasker (.align files). This allowed us to use the corrected percentage of divergence of each copy to the consensus from these .align files (accounting for the extremely high rate of mutations at CpG sites). In case of overlaps (when a position could be aligned to more than one consensus sequence), the smallest percentage divergence was chosen for that position.

DNA loss calculation

DNA loss coefficients were calculated as in [33]. We estimated lineage-specific DNA loss coefficients k with $E = A\,e^{-kt}$, where E is the amount of extant ancestral sequence in the species considered, A the total ancestral assembly size, and t the time, leading to k = ln(A/X)/t). Assuming, for Eutherians, A = 2,800 Mb and t = 100 My, we get k = 0.0026 My^{-1} for human (X = assembly size minus gains = 2,150 Mb). See Dataset S1 for all values and coefficients of other species. DNA loss rates in Mb/My were obtained by dividing the amount of loss by the divergence time.

Note that changing the ancestral assembly size would affect the numbers but not the differences between species or the correlations. For example, using 2,600 Mb as in the mouse genome paper [30] would simply reduce loss of 200 Mb for each mammal. For birds, there is not a reconstruction of the ancestral assembly size of Neoaves as there is for mammals [111]. The ancestral genome size of Neoaves is estimated to range from 1.5 to 1.7 Gb. Therefore, we considered low and high boundaries of ancestral assembly sizes being 1.2 and 1.4 Gb, respectively, and we used the middle point of 1.3 Gb for all calculations. We also characterized total loss at the same evolutionary timescales as the ones of our microdeletions and midsize deletion calculations (Figure S4).

Additionally, with this method, we only measure net totals of DNA gain and loss amounts. Therefore, they likely represent underestimations of genome dynamics, even more so for the species with higher TE removal. For example, two species may have similar measured gains and loss, but one would in fact have higher efficiency of TE removal. The latter would appear as lower gains, which consequently translates in lower loss with our method. Such bias can be addressed by measuring DNA gain and loss at shorter evolutionary time scales and verify the extent of the gain and loss dynamics (Figure S4). Importantly, the results at shorter time scales recapitulate the ones of the 70 My scale, with the woodpecker and

flightless bird (ostrich and penguins) lineages showing the most and the least dynamic genomes, respectively.

Analysis of microdeletions using a multi species alignment

We developed custom Perl scripts (MAFmicrodel), available at https://github.com/4ureliek/MAF_parsing). In a first step, we extracted alignment lines corresponding to species of interest (mammals or birds) from the MultiZ alignment of human chromosomes 1 to 22 with 100 other species from the UCSC genome browser (MAF format, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz100way/); see the usage of the MAF_microdel--1--get-gaps.pl script. Studied species are listed on Figure 5, with *Monodelphis domestica* and *Anolis carolinensis* as outgroups for mammals and birds, respectively (Dataset S2). Gaps were then extracted and listed in bed format for blocks containing alignment information for all species. For this analysis, we also filtered out blocks shorter than 50 bp (MAF_microdel--1--get-gaps.pl). In a second step (scripts MAF_microdel--2--analyze-gaps-XXX.pl), gaps in alignments were placed on a phylogenetic tree based on parsimony and using intersections of gap coordinates with Bedtools [112]. To limit biases arising from counting specific gaps occurring at the same location as shared, a gap is considered shared between two species only if the reciprocal overlap is >80% (options -f 0.80 and -r of intersectBed). The deletion rate is then calculated by dividing the amount of gaps strictly specific to each branch (not present in any of the other species) by the length of each branch (in My).

Analysis of microdeletions and midsize deletions for trio of species

We developed custom Perl scripts for the analysis of microdeletion and midsize deletions in species trios, available at https://github.com/4ureliek/DelGet. Most parameters can be adjusted through the configuration file. The steps are as follow: (i) Pick random positions in the outgroup species (ii) Extract X bp of sequences from this species' randomization, separated by Y bp. We chose X = 100 bp with Y = 10 kb (because the N50 of the contigs in some assemblies are <20 kb). (iii) Use BLAT to obtain hits in other species. Hits are kept only if (a) they are both on the same scaffold of the target; (b) both hit lengths are >120 bp; (c) both hits are on the same strand, score of best hit/next hit is >0.9 (to avoid uncertainty related to repeats), and (d) the region does not overlap with gaps in assembly ('N' nucleotides). (iv) The sequences between anchors are extracted for all three genomes. (v) These sequences are aligned with MUSCLE [113]. (vi) Gaps in alignments are placed on a

phylogenetic tree through intersections with Bedtools [112]. A gap is considered shared between 2 species if the reciprocal overlap is >85% (option -f 0.85 and option –r for reciprocity). Additionally, when only one base interrupts a gaps (e.g. GTGC---------A------ATGTC) it is skipped and the 2 gaps are merged in one (its length being corrected by 1 bp).

Screening for large deletions

Using a custom Perl script (maf_get_large_indels.pl, available at https://github.com/4ureliek/MAF_parsing), we recovered coordinates of indels >1 kb for each species in the MultiZ alignment of human chromosomes 1 to 22 with 100 other species from the UCSC genome browser (MAF format, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz100way/). The script outputs two types of large indels: first, a list of indels corresponding to cases where the sequence before and after is contiguous ("C" lines), implying that this region was either deleted in the source or inserted in the reference sequence (or a combination of both). This will lead to underestimating event lengths in case of shared repeats, but to avoid overestimating indel lengths due to TEs inserted in the reference (human), repeated sequences are discarded (as annotated by RepeatMasker and soft-masked in lower cases in the alignment). The script also outputs a list of indels corresponding to all consecutive empty data for a given species (no data on the browser, or double line when there are non aligning bases; these gaps could be shared between several species) only when the empty data is interrupting a scaffold in the species of interest. When there are inserted sequences in the species of interest, the script only outputs cases where the length in the reference (no lowercases) minus the insertion length is higher than the minimum length specified (here, 1 kb). Note that these gaps could arise from mis-assemblies or from segmental duplications in the reference, and all of them would require to be validated prior to any interpretations.

AKNOWLEDGMENTS

REFERENCES

1.   Petrov, D.A., *Evolution of genome size: new approaches to an old problem.* Trends in Genetics, 2001. **17**(1): p. 23-28.

2.   Gregory, T.R., *The Evolution of the Genome*, ed. T.R. Gregory. 2005, San Diego, CA: Elsevier Academic Press.

3.   Lynch, M., *The origins of genome architecture*. 2007, Sunderland, MA.: Sinauer Assocs., Inc.

4.   Linquist, S., et al., *Applying ecological models to communities of genetic elements: the case of neutral theory.* Molecular ecology, 2015. **24**(13): p. 3232-3242.

5.   Canapa, A., et al., *Transposons, Genome Size, and Evolutionary Insights in Animals.* Cytogenetic and genome research, 2016. **147**(4): p. 217-39.

6.   Cavalier-Smith, T., *Skeletal DNA and the evolution of genome size.* Annual review of biophysics and bioengineering, 1982. **11**: p. 273-302.

7.   Gregory, T.R. and P.D. Hebert, *The modulation of DNA content: proximate causes and ultimate consequences.* Genome Research, 1999. **9**(4): p. 317-324.

8.   Andrews, C.B., S.A. Mackenzie, and T.R. Gregory, *Genome size and wing parameters in passerine birds.* Proceedings. Biological sciences / The Royal Society, 2009. **276**(1654): p. 55-61.

9.   Wright, N.A., T.R. Gregory, and C.C. Witt, *Metabolic 'engines' of flight drive genome size reduction in birds.* Proceedings. Biological sciences / The Royal Society, 2014. **281**(1779): p. 20132780.

10.   Petrov, D.A., *Mutational Equilibrium Model of Genome Size Evolution.* Theoretical population biology, 2002. **61**(4): p. 531-44.

11.   Nam, K. and H. Ellegren, *Recombination drives vertebrate genome contraction.* PLoS genetics, 2012. **8**(5): p. e1002680.

12.   Vicient, C.M., et al., *Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus Hordeum.* Plant Cell, 1999. **11**(9): p. 1769-1784.

13.   Bennetzen, J.L., J. Ma, and K.M. Devos, *Mechanisms of recent genome size variation in flowering plants.* Annals of botany, 2005. **95**(1): p. 127-132.

14.   Piegu, B., et al., *Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice.* Genome Research, 2006. **16**(10): p. 1262-1269.

15.   Vitte, C., O. Panaud, and H. Quesneville, *LTR retrotransposons in rice (Oryza sativa, L.): recent burst amplifications followed by rapid DNA loss.* BMC genomics, 2007. **8**: p. 218.

16.   Hawkins, J.S., et al., *Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants.* Proceedings of the National Academy of Sciences, 2009. **106**(42): p. 17811-17816.

17. Kelly, L.J., et al., *Analysis of the giant genomes of Fritillaria (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size.* The New phytologist, 2015. **208**(2): p. 596-607.

18. Nystedt, B., et al., *The Norway spruce genome sequence and conifer genome evolution.* Nature, 2013. **497**(7451): p. 579-584.

19. Petrov, D.A., E.R. Lozovskaya, and D.L. Hartl, *High intrinsic rate of DNA loss in Drosophila.* Nature, 1996. **384**(6607): p. 346-349.

20. Petrov, D.A., et al., *Evidence for DNA loss as a determinant of genome size.* Science, 2000. **287**(5455): p. 1060-1062.

21. Bensasson, D., et al., *Genomic gigantism: DNA loss is slow in mountain grasshoppers.* Molecular biology and evolution, 2001. **18**(2): p. 246-253.

22. Wang, X., et al., *The locust genome provides insight into swarm formation and long-distance flight.* Nature communications, 2014. **5**: p. 2957.

23. Blass, E., M. Bell, and S. Boissinot, *Accumulation and rapid decay of non-LTR retrotransposons in the genome of the three-spine stickleback.* Genome biology and evolution, 2012. **4**(5): p. 687-702.

24. Aparicio, S., et al., *Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.* Science, 2002. **297**(5585): p. 1301-1310.

25. Neafsey, D.E. and S.R. Palumbi, *Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes.* Genome Research, 2003. **13**(5): p. 821-830.

26. Sun, C., J.R. López Arriaza, and R.L. Mueller, *Slow DNA loss in the gigantic genomes of salamanders.* Genome biology and evolution, 2012. **4**(12): p. 1340-1348.

27. Sun, C., et al., *LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders.* Genome biology and evolution, 2012. **4**(2): p. 168-183.

28. Schubert, I. and G.T.H. Vu, *Genome Stability and Evolution: Attempting a Holistic View.* Trends in plant science, 2016. **21**(9): p. 749–757.

29. Van de Peer, Y., S. Maere, and A. Meyer, *The evolutionary significance of ancient genome duplications.* Nature reviews. Genetics, 2009. **10**(10): p. 725-732.

30. Consortium, M.G.S., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

31. Gibbs, R.A., et al., *Genome sequence of the Brown Norway rat yields insights into mammalian evolution.* Nature, 2004. **428**(6982): p. 493-521.

32. Hillier, L.W., et al., *Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.* Nature, 2004. **432**(7018): p. 695-716.

33. Lindblad-Toh, K., et al., *Genome sequence, comparative analysis and haplotype structure of the domestic dog.* Nature, 2005. **438**(7069): p. 803-819.

34. Consortium, R.M.G.S.a.A., et al., *Evolutionary and biomedical insights from the rhesus macaque genome.* Science, 2007. **316**(5822): p. 222-234.

35. She, X., et al., *Mouse segmental duplication and copy number variation.* Nature genetics, 2008. **40**(7): p. 909-914.

36. John, J.S. and T.W. Quinn, *Identification of novel CR1 subfamilies in an avian order with recently active elements.* Molecular phylogenetics and evolution, 2008. **49**(3): p. 1008-14.

37.   Nicholas, T.J., et al., *The genomic architecture of segmental duplications and associated copy number variants in dogs.* Genome Research, 2009. **19**(3): p. 491-499.
38.   Consortium, B.G.S.a.A., et al., *The genome sequence of taurine cattle: a window to ruminant biology and evolution.* Science, 2009. **324**(5926): p. 522-528.
39.   Bolisetty, M., et al., *Unexpected Diversity and Expression of Avian Endogenous Retroviruses.* mBio, 2012. **3**(5): p. e00344-12.
40.   Zhang, G., et al., *Comparative genomics reveals insights into avian genome evolution and adaptation.* Science, 2014. **346**(6215): p. 1311-1320.
41.   Chalopin, D., et al., *Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates.* Genome biology and evolution, 2015. **7**(2): p. 567-580.
42.   Ji, Y. and J.A. DeWoody, *Genomic Landscape of Long Terminal Repeat Retrotransposons (LTR-RTs) and Solo LTRs as Shaped by Ectopic Recombination in Chicken and Zebra Finch.* Journal of molecular evolution, 2016. **82**(6): p. 251-63.
43.   Thomas, J.W., et al., *Comparative analyses of multi-species sequences from targeted genomic regions.* Nature, 2003. **424**(6950): p. 788-793.
44.   Organ, C.L., et al., *Origin of avian genome size and structure in non-avian dinosaurs.* Nature, 2007. **446**(7132): p. 180-184.
45.   Hughes, A.L. and M.K. Hughes, *Small genomes for better flyers.* Nature, 1995. **377**(6548): p. 391.
46.   Zhang, Q. and S.V. Edwards, *The evolution of intron size in amniotes: a role for powered flight?* Genome biology and evolution, 2012. **4**(10): p. 1033-1043.
47.   Lindblad-Toh, K., et al., *A high-resolution map of human evolutionary constraint using 29 mammals.* Nature, 2011. **478**(7370): p. 476-482.
48.   Smit, A.F.A., R. Hubley, and P. Green, *RepeatMasker Open-4.0.2013-2015* http://www.repeatmasker.org.
49.   Jarvis, E.D., et al., *Whole-genome analyses resolve early branches in the tree of life of modern birds.* Science, 2014. **346**(6215): p. 1320-1331.
50.   Blanchette, M., et al., *Reconstructing large regions of an ancestral mammalian genome in silico.* Genome Research, 2004. **14**(12): p. 2412-2423.
51.   Ma, J., et al., *Reconstructing contiguous regions of an ancestral genome.* Genome Research, 2006. **16**: p. 1557-1565.
52.   Organ, C.L. and S.V. Edwards, *Major events in avian genome evolution* in *Living Dinosaurs: The Evolutionary History of Modern Birds*, D.G.K. G, Editor. 2011, John Wiley & Sons, Ltd. p. 325-337.
53.   Ophir, R. and D. Graur, *Patterns and rates of indel evolution in processed pseudogenes from humans and murids.* Gene, 1997. **205**(1-2): p. 191-202.
54.   Laurie, S., et al., *Sequence shortening in the rodent ancestor.* Genome Research, 2012. **22**(3): p. 478-485.
55.   Bromham, L., *The genome as a life-history character: why rate of molecular evolution varies between mammal species.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2011. **366**(1577): p. 2503-2513.
56.   Sung, W., et al., *Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life.* G3, 2016. **6**(8): p. 2583-2591.

57.   Shedlock, A.M., *Phylogenomic investigation of CR1 LINE diversity in reptiles.* Syst Biol, 2006. **55**(6): p. 902-11.
58.   Shedlock, A.M., et al., *Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome.* Proc Natl Acad Sci U S A, 2007. **104**(8): p. 2767-72.
59.   Janes, D.E., et al., *Genome evolution in Reptilia, the sister group of mammals.* Annual review of genomics and human genetics, 2010. **11**: p. 239-264.
60.   Adams, M.D., et al., *The genome sequence of Drosophila melanogaster.* Science, 2000. **287**(5461): p. 2185-2195.
61.   Gregory, T.R., *Is small indel bias a determinant of genome size?* Trends in genetics, 2003. **19**(9): p. 485-488.
62.   Gregory, T.R., *Insertion-deletion biases and the evolution of genome size.* Gene, 2004. **324**: p. 15-34.
63.   Mills, R.E., *An initial map of insertion and deletion (INDEL) variation in the human genome.* Genome Research, 2006. **16**(9): p. 1182-1190.
64.   Nobrega, M.A., et al., *Megabase deletions of gene deserts result in viable mice.* Nature, 2004. **431**(7011): p. 988-993.
65.   Han, K., et al., *Alu recombination-mediated structural deletions in the chimpanzee genome.* PLoS genetics, 2007. **3**(10): p. 1939-1949.
66.   McLean, C. and G. Bejerano, *Dispensability of mammalian DNA.* Genome Research, 2008. **18**(11): p. 1743-1751.
67.   Ponting, C.P., C. Nellaker, and S. Meader, *Rapid turnover of functional sequence in human and other genomes.* Annu Rev Genomics Hum Genet, 2011. **12**: p. 275-99.
68.   Rands, C.M., et al., *8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage.* PLoS Genet, 2014. **10**(7): p. e1004525.
69.   Chan, Y.F., et al., *Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer.* Science, 2010. **327**(5963): p. 302-5.
70.   McLean, C.Y., et al., *Human-specific loss of regulatory DNA and the evolution of human-specific traits.* Nature, 2011. **471**(7337): p. 216-9.
71.   Hiller, M., B.T. Schaar, and G. Bejerano, *Hundreds of conserved non-coding genomic regions are independently lost in mammals.* Nucleic Acids Res, 2012. **40**(22): p. 11463-76.
72.   Meredith, R.W., et al., *Evidence for a single loss of mineralized teeth in the common avian ancestor.* Science, 2014. **346**(6215): p. 1254390.
73.   Lovell, P.V., et al., *Conserved syntenic clusters of protein coding genes are missing in birds.* Genome biology, 2014. **15**(12): p. 565.
74.   Chaisson, M.J., R.K. Wilson, and E.E. Eichler, *Genetic variation and the de novo assembly of human genomes.* Nat Rev Genet, 2015. **16**(11): p. 627-40.
75.   Ellegren, H. and N. Galtier, *Determinants of genetic diversity.* Nature Reviews Genetics, 2016. **17**: p. 422-433.
76.   Lynch, M., et al., *The repatterning of eukaryotic genomes by random genetic drift.* Annu Rev Genomics Hum Genet, 2011. **12**: p. 347-66.
77.   Whitney, K.D., et al., *Drift and genome complexity revisited.* PLoS Genet, 2011. **7**(6): p. e1002092.

78.   Figuet, E., et al., *Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes.* Molecular Miology and Evolution, 2016. **33**(6): p. 1517-1527.

79.   Szitenberg, A., et al., *Genetic drift, not life history or RNAi, determine long term evolution of transposable elements.* Genome Biology and Evolution, 2016. **8**(9): p. 2964-2978.

80.   Konkel, M.K. and M.A. Batzer, *A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome.* Seminars in cancer biology, 2010. **20**(4): p. 211-221.

81.   Shirasu, K., et al., *A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion.* Genome Research, 2000. **10**(7): p. 908-15.

82.   Devos, K.M., J.K.M. Brown, and J.L. Bennetzen, *Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis.* Genome Research, 2002. **12**(7): p. 1075-1079.

83.   Ma, J., K.M. Devos, and J.L. Bennetzen, *Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice.* Genome Research, 2004. **14**(5): p. 860-869.

84.   Tiley, G.P. and J.G. Burleigh, *The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms.* BMC Evol Biol, 2015. **15**: p. 194.

85.   van de Lagemaat, L.N., et al., *Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates.* Genome Research, 2005. **15**(9): p. 1243-9.

86.   Sen, S.K., et al., *Human genomic deletions mediated by recombination between Alu elements.* American journal of human genetics, 2006. **79**(1): p. 41-53.

87.   Startek, M., et al., *Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination.* Nucleic Acids Research, 2015. **43**(4): p. 2188-98.

88.   Lee, J., et al., *Human Genomic Deletions Generated by SVA-Associated Events.* Comparative and functional genomics, 2012. **2012**: p. 807270.

89.   Pagán, H.J.T., et al., *Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats.* Genome Biology and Evolution, 2012. **4**(4): p. 575-585.

90.   Platt, R.N., et al., *Targeted capture of phylogenetically-informative Ves SINE insertions in genus Myotis.* Genome Biology and Evolution, 2015. **7**(6): p. 1664-1675.

91.   Organ, C.L. and A.M. Shedlock, *Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates.* Biology letters, 2009. **5**(1): p. 47-50.

92.   Gregory, T.R., et al., *The smallest avian genomes are found in hummingbirds.* Proceedings. Biological sciences / The Royal Society, 2009. **276**(1674): p. 3753-3757.

93.   Shen, Y.-Y., et al., *Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability.* Genome Research, 2009. **19**(10): p. 1760-1765.

94.   Gregory, T.R., *Genome size and developmental parameters in the homeothermic vertebrates.* Genome, 2002. **45**(5): p. 833-838.

95.   Clarke, A., P. Rothery, and N.J. Isaac, *Scaling of basal metabolic rate with body mass and temperature in mammals.* The Journal of animal ecology, 2010. **79**(3): p. 610-9.

96.   Vinogradov, A.E., *Nucleotypic Effect in Homeotherms: Body-Mass Independent Resting Metabolic Rate of Passerine Birds is Related to Genome Size.* Evolution, 1997. **51**(1): p. 220-226.

97.   Smith, J.D.L. and T.R. Gregory, *The genome sizes of megabats (Chiroptera: Pteropodidae) are remarkably constrained.* Biology letters, 2009. **5**(3): p. 347-351.

98.   Garland, T., *The Relation between Maximal Running Speed and Body-Mass in Terrestrial Mammals.* Journal of Zoology, 1983. **199**(Feb): p. 157-170.

99.   Stuart, J.A. and M.M. Page, *Plasma IGF-1 is negatively correlated with body mass in a comparison of 36 mammalian species.* Mech Ageing Dev, 2010. **131**(9): p. 591-8.

100.   Weber, C.C., et al., *Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection.* Genome Biology, 2014. **15**(12): p. 542.

101.   Bao, W., K.K. Kojima, and O. Kohany, *Repbase Update, a database of repetitive elements in eukaryotic genomes.* Mob DNA, 2015. **6**: p. 11.

102.   Pace, J.K., et al., *Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods.* Proceedings of the National Academy of Sciences, 2008. **105**(44): p. 17023-17028.

103.   Ray, D.A., et al., *Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus.* Genome Research, 2008. **18**(5): p. 717-728.

104.   Mitra, R., et al., *Functional characterization of piggyBat from the bat Myotis lucifugus unveils an active mammalian DNA transposon.* Proceedings of the National Academy of Sciences, 2013. **110**(1): p. 234-239.

105.   Thomas, J., et al., *Rolling-circle transposons catalyze genomic innovation in a mammalian lineage.* Genome biology and evolution, 2014. **6**(10): p. 2595-2610.

106.   Zhuo, X. and C. Feschotte, *Cross-Species Transmission and Differential Fate of an Endogenous Retrovirus in Three Mammal Lineages.* PLoS pathogens, 2015. **11**(11): p. e1005279.

107.   Eory, L., et al., *Avianbase: a community resource for bird genomics.* Genome Biol, 2015. **16**: p. 21.

108.   Warren, W.C., et al., *The genome of a songbird.* Nature, 2010. **464**(7289): p. 757-762.

109.   Suh, A., et al., *Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds.* Nature communications, 2011. **2**.

110.   Suh, A., et al., *Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes.* Genome biology and evolution, 2015. **7**(1): p. 205-217.

111.   Green, R.E., et al., *Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs.* Science, 2014. **346**(6215): p. 1254449.

112.   Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics (Oxford, England), 2010. **26**(6): p. 841-842.

113.   Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.

114. Gregory, T.R., *Animal Genome Size Database.* http://www.genomesize.com. 2016.
115. Shedlock, A.M. and S.V. Edwards, *Amniotes (Amniota)*, in *The Timetree of Life*, S.B.a.K. Hedges, S., Editor. 2009, Oxford University Press. p. 375–379.
116. Meredith, R.W., et al., *Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification.* Science, 2011. **334**(6055): p. 521-4.
117. Team, R.D.C., *R: A language and environment for statistical computing.* 2008, R Foundation for Statistical Computing: Vienna, Austria.
118. Teeling, E.C., et al., *A molecular phylogeny for bats illuminates biogeography and the fossil record.* Science, 2005. **307**(5709): p. 580-4.
119. Hedges, S.B., J. Dudley, and S. Kumar, *TimeTree: a public knowledge-base of divergence times among organisms.* Bioinformatics, 2006. **22**(23): p. 2971-2.

FIGURE LEGENDS

Figure 1: Genome size variation in amniotes

Genome size ranges of different groups of species are shown as black bars (from smallest to largest genome sizes). Birds range from ~1 to 2.1 Gb and mammals from 1.6 to 6.3 Gb, which is in contrast to the range among all vertebrates (~0.4 to ~133 Gb) [114]. Divergence times are represented on the phylogenetic tree in My as in [115, 116]. Red dot: median. N: numbers of species inside each group with genome size data included in the figure (compiled from [114] as of March 6th 2015). When several measures exist for one single species, values are averaged. Gb: gigabases. For rodents, the red viscacha rat was not included (tetraploid, genome size estimated to 8.4 Gb).

Figure 2: Gain and loss of DNA in 10 mammalian lineages

For each species, phylogenetic relationship (left panel) [116], TE content (light blue bars), assembly sizes (with N removed, grey bars), DNA loss coefficients (green bars) as well as gain (red and orange) and loss (dark blue) of DNA are shown. DNA gains correspond mostly to lineage-specific TEs in red (not shared with other mammals). When available, measures of segmental duplications (SDs) were added (orange). Because SDs also contain TEs, we corrected the SD amounts with the TE content of each genomes. DNA loss amounts and rates are calculated as in [33] using a common ancestor 'assembly' size for all mammals of 2,800 Mb (Methods). Phylogenetic tree is color-coded based on genome sizes in pg (combination of data from [114]) based on parsimony. See Dataset S1 for numbers, calculation steps and assemblies details. All numbers are in Mb.

Figure 3: Gain and loss of DNA in 12 avian lineages

For each species, phylogenetic relationship (left panel) [49], TE content (light blue bars), assembly sizes (with N removed, grey bars), DNA loss coefficients (green bars) as well as gain (red) and loss (dark blue) of DNA are shown. Species names in bold correspond to high coverage genomes, and the others to low coverage genomes [40]. DNA gain corresponds to insertions younger than 70 My. DNA loss amounts and rates are calculated as in [33] using a common ancestor size of 1,300 Mb (see text and Methods). Phylogenetic tree is color coded based on genome sizes in pg (combination of data from [114] and extrapolations from assembly sizes and coverage), based on parsimony as in [52]. See Dataset S1 for numbers, calculation steps and assemblies details. All numbers are in Mb.

Figure 4: Gain and loss of DNA as driving forces of genome size variation

DNA gains are plotted in function of DNA loss coefficients for the 24 birds examined. With $R^2 = 0.658$ (r^2 on graph) and n = 24, $p = 0.000474$. The values and R [117] command lines used to build this figure can be found in Dataset S1.

Figure 5: Microdeletion rates across amniotes

Microdeletion (1-30 bp) rates and number of events are shown in green and purple respectively. Microdeletions are estimated from gaps in the UCSC MultiZ 100-species alignment of human chromosomes 1 to 22, restricted to blocks containing information for the species studied (Methods and Dataset S2). Deletion rates are calculated by dividing the amount of gaps specific to each branch (not present in any other species) by My of each branch. My: million years. Gb: gigabases. Genome sizes are from [114]. There were no reported genome sizes for three species, so the size of the closest species is shown: the Tenrecidae *Setifer setosus* for *E. telfari*, *M. fuscata* for *M. mulatta*, and the average of two birds of the same family (Emberizidae) for the medium ground finch. Scales are indicated on top (note the difference between panels A and B). A. Microdeletion rates in 11 placental mammals (with *Monodelphis domestica* as outgroup). The total length of the alignment is 297 Mb and timescales are as in [31, 116, 118, 119]. Names of orders are indicated on the tree (in blue). B. Microdeletion rates in seven birds (with *Anolis carolinensis* as outgroup). The total length of the alignment is 66.5 Mb. Timescales are from [49]. Names of two super-orders are placed on the tree (in blue).

Figure 6: Midsize deletion rates across amniotes

Microdeletions (1 to 30 bp) and midsize deletion (from 30 bp to 10 kb for mammals and birds, respectively) rates, measured after a recent split between two species, are shown in light green and dark blue, respectively. Deletion rates were calculated based on gaps in alignments of orthologous specific to the species (Methods) and are normalized by alignment length (in Mb on the figure). We placed the deletion events on the phylogenetic tree based on parsimonious polarization via the respective outgroup species. Gaps were considered shared by multiple species when they overlapped for at least 85% of their respective lengths (Methods). For bats, 2 sets of 3 species were considered: group I with *M. lucifugus*, *M. brandtii*, and *E. fuscus* (113.1 Mb of alignment), group II with *M. lucifugus*, *E. fuscus*, and *P. vampyrus* (97.9 Mb of alignment). Rates placed on the *Myotis* branch (after the split with *E. fuscus* but before the split between *M. lucifugus* and *M. brandtii*) were inferred from *M. lucifugus* rates. See Dataset S3 for the numbers, and Figure S3 for the midsize deletion spectrums.

FIGURES
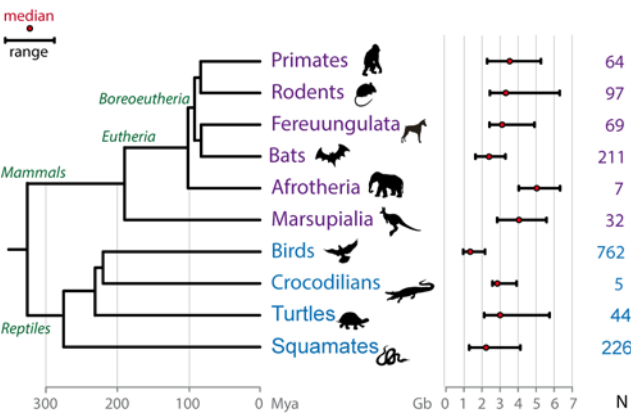
Figure 1: Genome size variation in amniotes
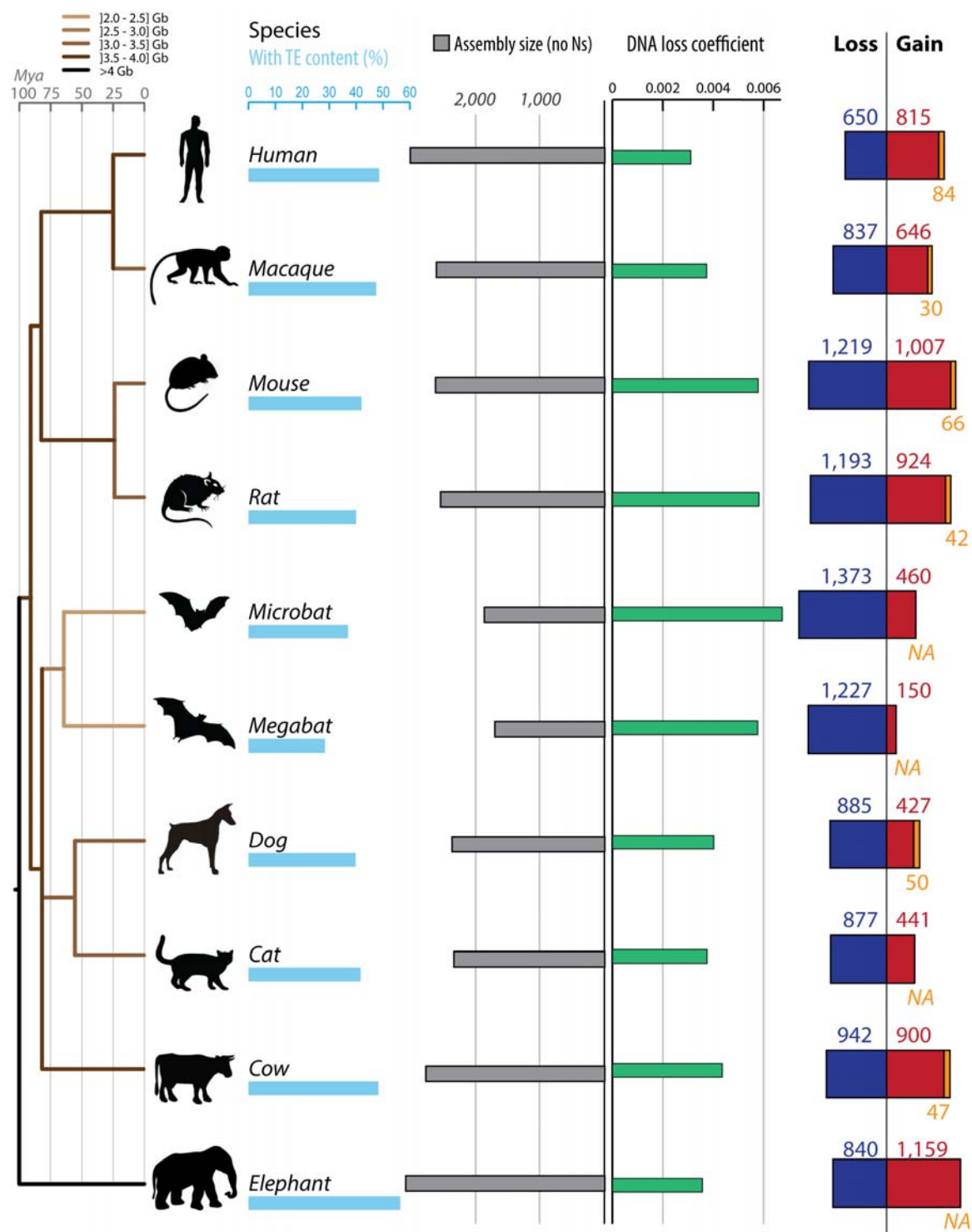
Figure 2: Gain and loss of DNA in 10 mammalian lineages
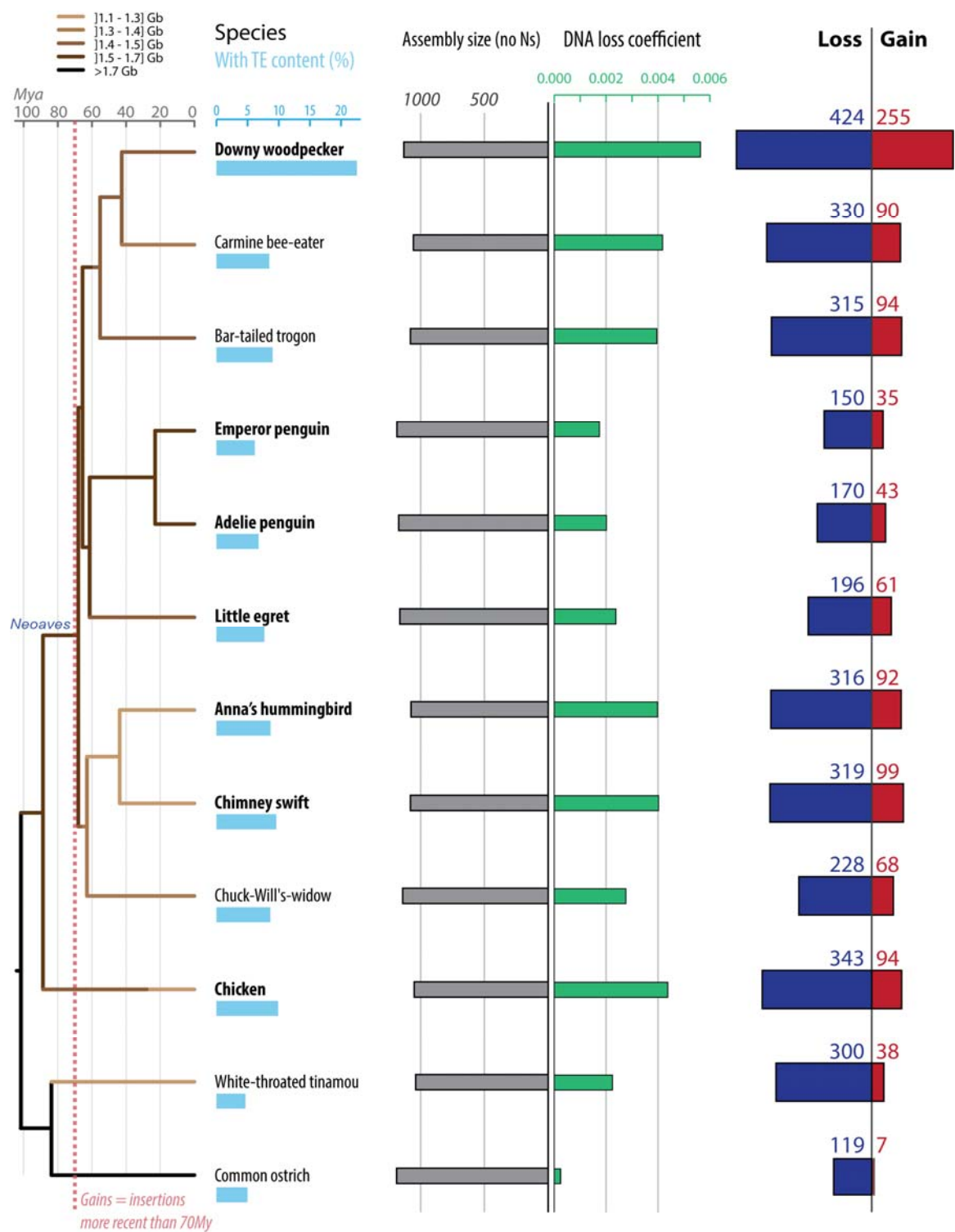
Figure 3: Gain and loss of DNA in 12 avian lineages
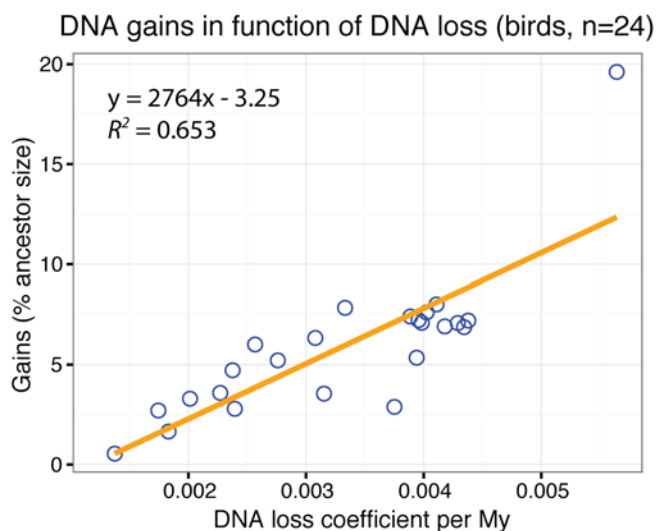
Figure 4: Gain and loss of DNA as driving forces of genome size variation
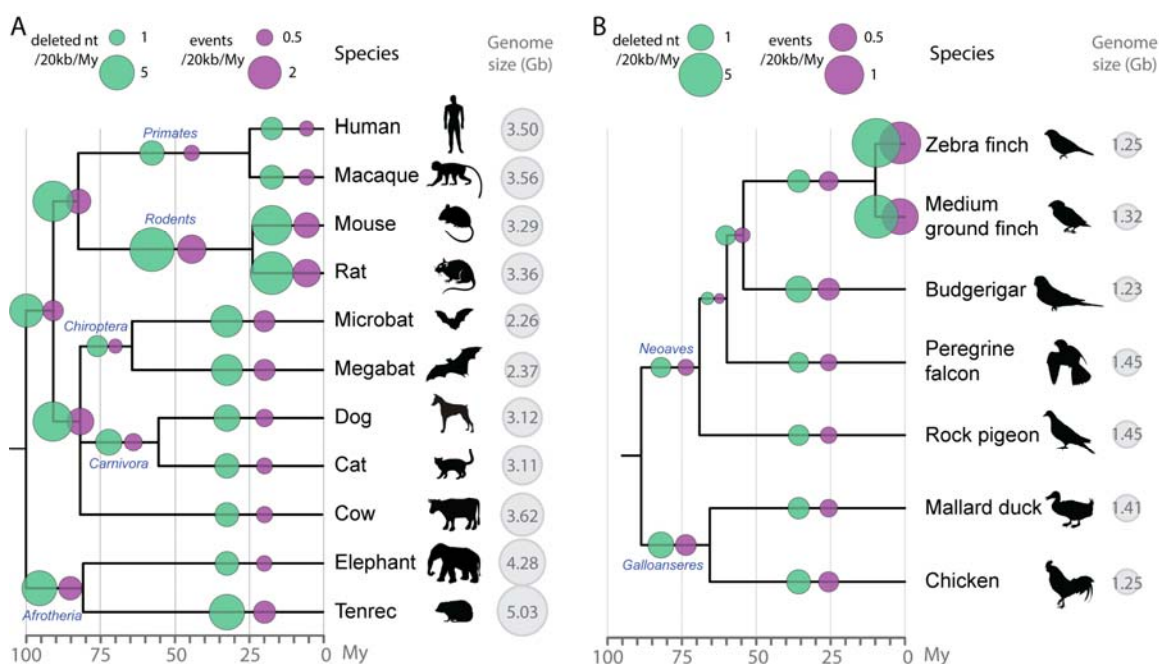
Figure 5: Microdeletion rates across amniotes
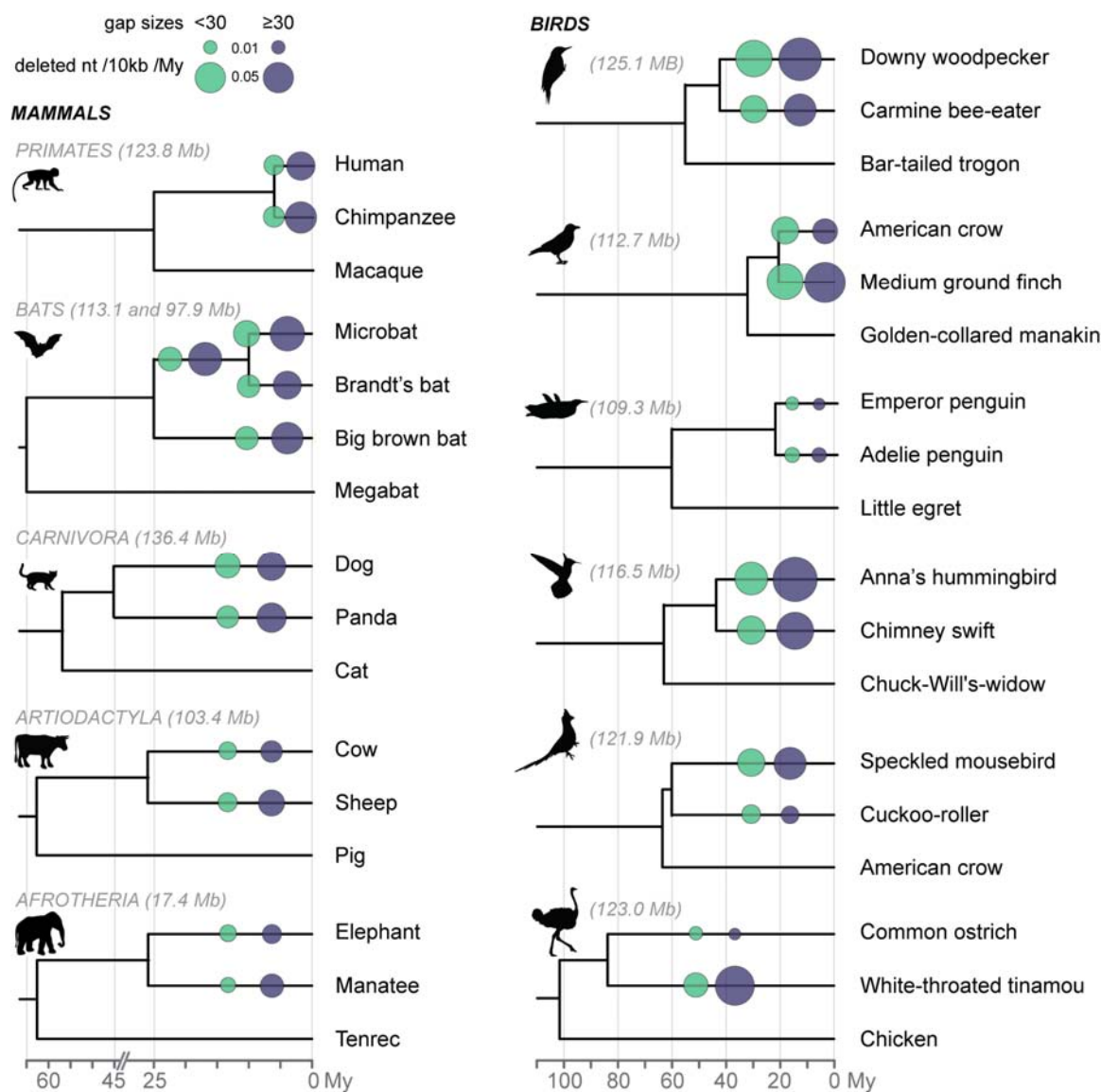
Figure 6: Midsize deletion rates across amniotes

SUPPLEMENTARY MATERIAL

Dataset S1: gain and loss of DNA (data of Figures 2, 3, 4 S1 and S4)

Dataset S2: Analysis of microdeletions using a multi species alignment (data of Figure 5)

Dataset S3: Analysis of microdeletions and midsize deletions for trio of species (data of Figure 6)

Dataset S4: Statistical significance for the pairwise comparisons of distributions of Figures S2, S3 and S5

Figure S1

Species	Loss	Gain
Medium ground finch	341	90
American crow	337	92
Golden-collared manakin	252	82
Budgerigar	310	96
Kea	313	69
Peregrine falcon	191	47
Speckled mousebird	325	104
White-tailed eagle	201	36
Turkey vulture	156	22
MacQueen's bustard	257	46
Red-crested turaco	214	78
Common cuckoo	271	102

Legend:
-]1.1 - 1.3] Gb
-]1.3 - 1.4] Gb
-]1.4 - 1.5] Gb
-]1.5 - 1.7] Gb
- >1.7 Gb

Species With TE content (%)

Assembly size (no Ns)   DNA loss rate per My

Gains = insertions more recent than 70My

Figure S1: Gain and loss of DNA in 12 additional birds

Figure reads exactly as Figure 3, with different species represented. For each species, phylogenetic relationship (left panel) [49], TE content (light blue bars), assembly sizes (with N removed, grey bars), DNA loss rates per My (green bars) as well as gain (red) and loss (dark blue) of DNA are shown. Species names in bold correspond to high-coverage genomes, and the others to low-coverage genomes [40]. DNA gain corresponds to TE insertions younger than 70 My. DNA loss amounts and rates are calculated as in [33] using a common ancestor size of 1,300 Mb (Methods). Phylogenetic tree is color coded based on genome sizes in pg (combination of data from [114] and extrapolations from assembly sizes and coverage), based on parsimony as in [52]. See Dataset S1 for numbers, calculation steps and assemblies details. All numbers are in Mb.
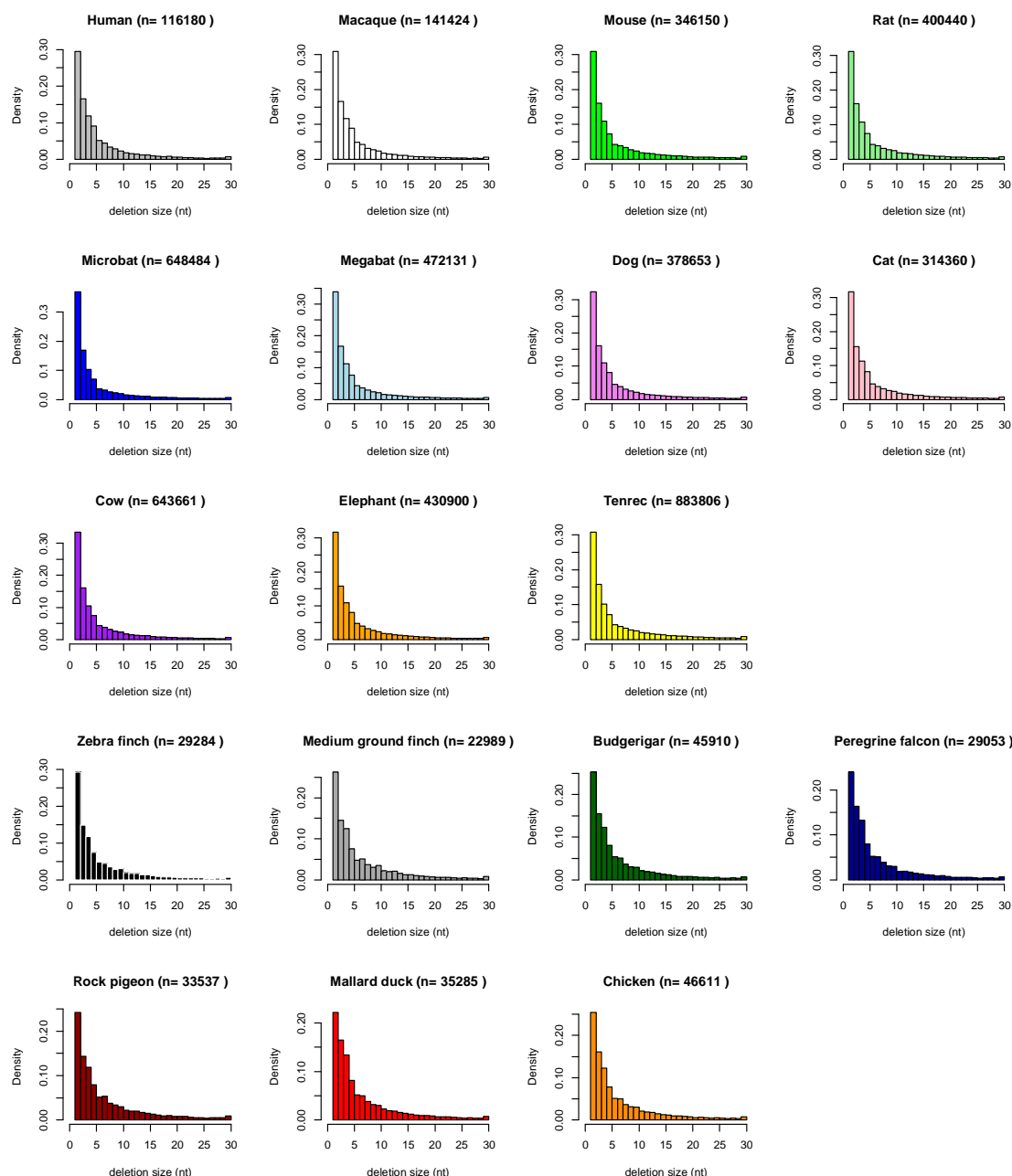
Figure S2: Microdeletion spectrums of mammals and birds
The data plotted is that of Figure 5, for the terminal branches (gaps specific to each species). The histograms are done in R [117], with for example for human: hist(hs$V1, freq=F, right=F, border="black", col="grey", main=paste("Human (n=",length(hs$V1),")"), xlab = "deletion size (nt)", xlim=c(0,30),breaks=30), xlim=c(0,30), breaks=30)), where hs corresponds to the list of gap lengths specific to human. Statistical significance is based on 1,000 replicates of Kolmogorov–Smirnov tests on samples of 5,000 values: two distributions are considered significantly different when at least 950 out of 1,000 tests have $p < 0.05$. Data and R command lines can be found in Dataset S4, sheet 'micro_del'. Densities and not frequencies are plotted to allow comparison between species. The number of deletion events (inferred from gaps in the alignment) is listed on top of the histograms (n).

Figure S3A
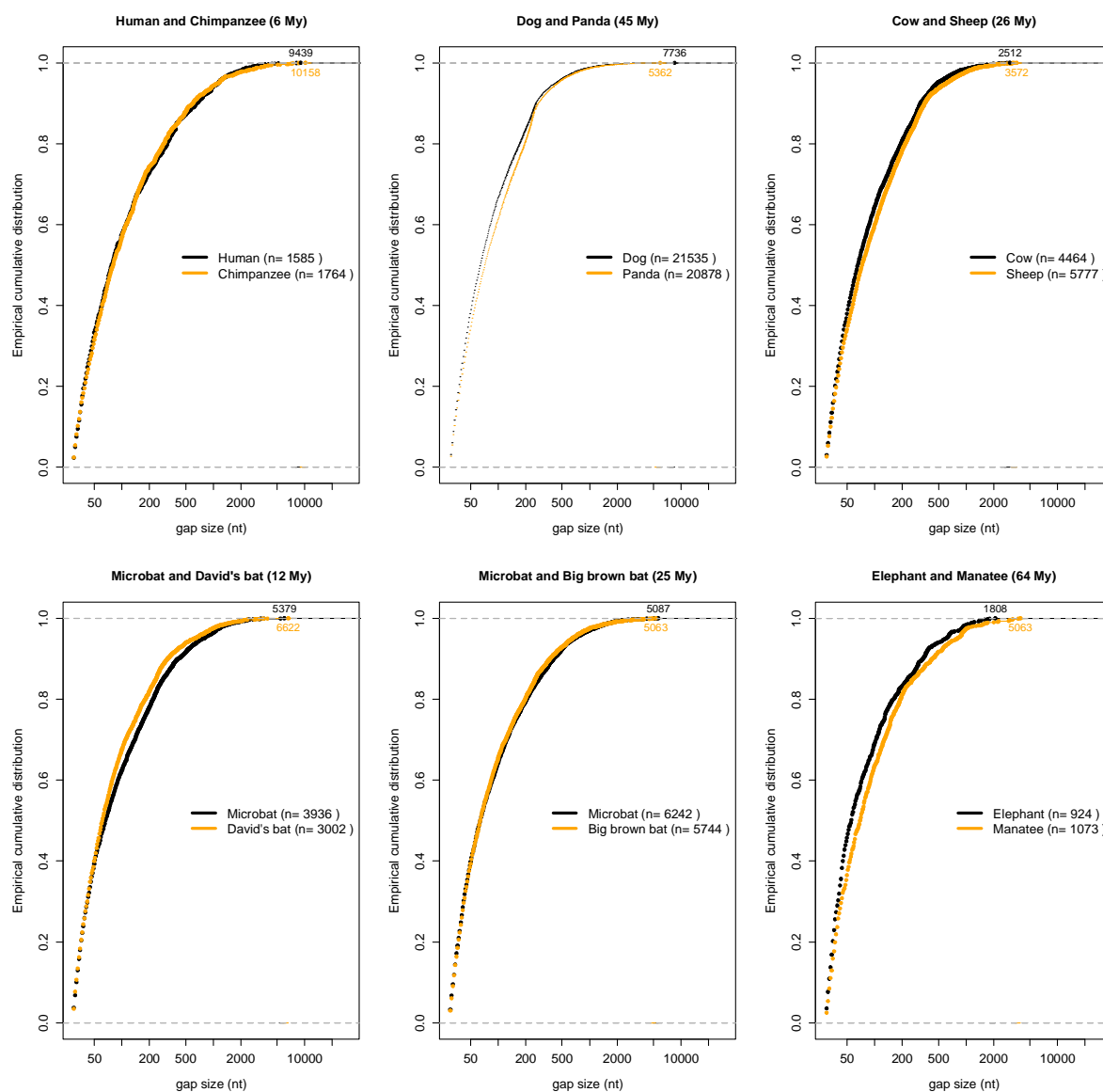
Figure S3B
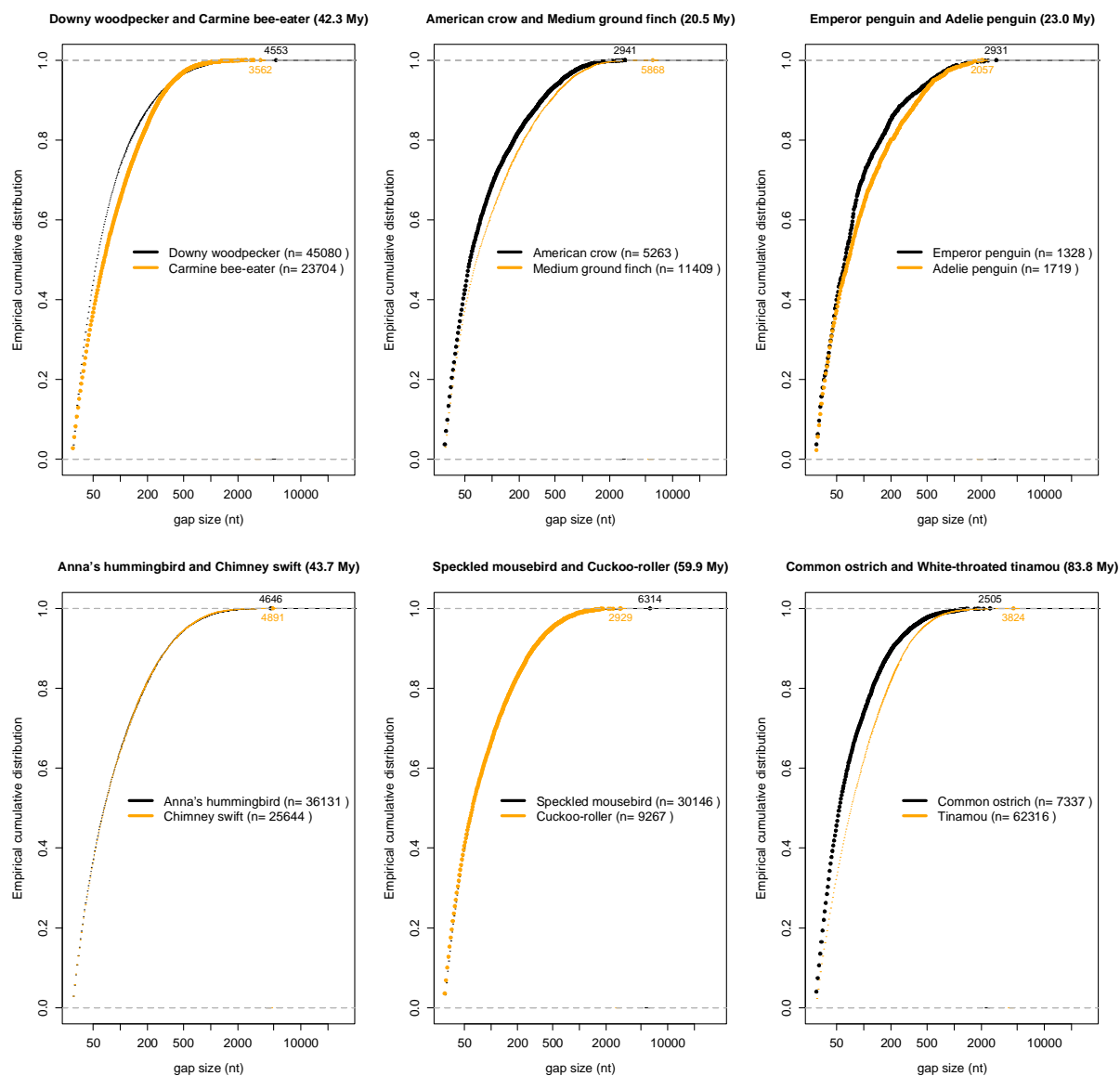
Figure S3C
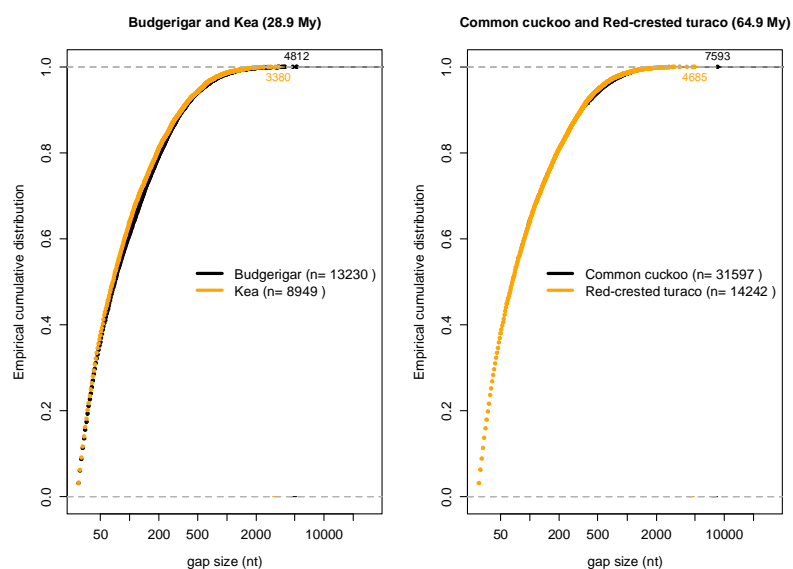
Figure S3: Midsize deletion spectrums of mammals and birds

A. Mammals (data of Figure 6). B. Birds (data of Figure 6). C. Four additional birds. The empirical cumulative distributions (ecd) are done in R [117] (command lines can be found in Dataset S3), and statistical tests are done as in Figure S2 (detailed in Dataset S4). Note the log scale for the x axis, for better visualization. On ecd graphs, differences in the slopes of the curves reveal differences in density: for example, the elephant has more deletions of ~200 bp than the manatee. Indeed, such representation (cumulative) allows to directly visualize the proportion of events under a certain size; for example, between 10 and 23% of the deletion events recovered in each species were larger than 200 bp. Clearly distinct curves suggest significant differences, but the dog and the panda distributions are the only ones that have enough events to be found significantly different (Dataset S4). For C (four additional birds), the DelGet pipeline (Methods) was run for the budgerigar and kea with the American crow as outgroup (155.1 Mb of alignment), and for the common cuckoo and the red-crested turaco with the Chuck-Will's-widow as outgroup (128.5 Mb of alignment). Resulting microdeletion and midsize deletion rates are intermediate to those of ostrich and Anna's hummingbird (Dataset S3). The largest events are labeled on the graphs and have been validated *in silico* through careful manual examination of the alignments. Additionally, their length has been corrected in case of lineage-specific TE insertions in the other species. Breakpoints are as follow for mammals: human (hg38 chr10:81816824), chimpanzee (panTro4 chr14:92050068), dog (canFam3 chr20:33042110), panda (ailMel1 GL194157.1:120088), microbat (myoLuc2 GL429820:6762144 for group I and GL429772:17082212 for group II), Brandt's bat (ASM41265v1 gb|KE161863.1|:221999), big brown bat (Eptesicus_fuscus_assembly1.0 gb|ALEH01017523.1|:41408), cow (bosTau8 chr22:10255818), sheep (oviAri3 chr11:31801480), manatee (TriManLat1.0 gi|460713488|ref|NW_004444110.1|:1177735), elephant (loxAfr3 scaffold_6:35237920). For birds, assemblies are as in [40] (Dataset S1) and breakpoints are as follow: downy woodpecker (scaffold783:2013016), carmine bee-eater (scaffold45549:18816), American crow (scaffold116:6801125), medium ground finch (scaffold40:8463065), emperor penguin (Scaffold500:719123), Adelie penguin (Scaffold241:3188990), Anna's hummingbird (scaffold104:965959), chimney swift (scaffold114:243360), speckled mousebird (scaffold41959:22764), cuckoo-roller (scaffold18151:34945), common ostrich (scaffold638:67481), tinamou (scaffold12594:11743), budgerigar (scf900160277013:1211857), kea (scaffold13226:14568), common cuckoo (scaffold492:288210), red-crested turaco (scaffold3767:111605).

Figure S4
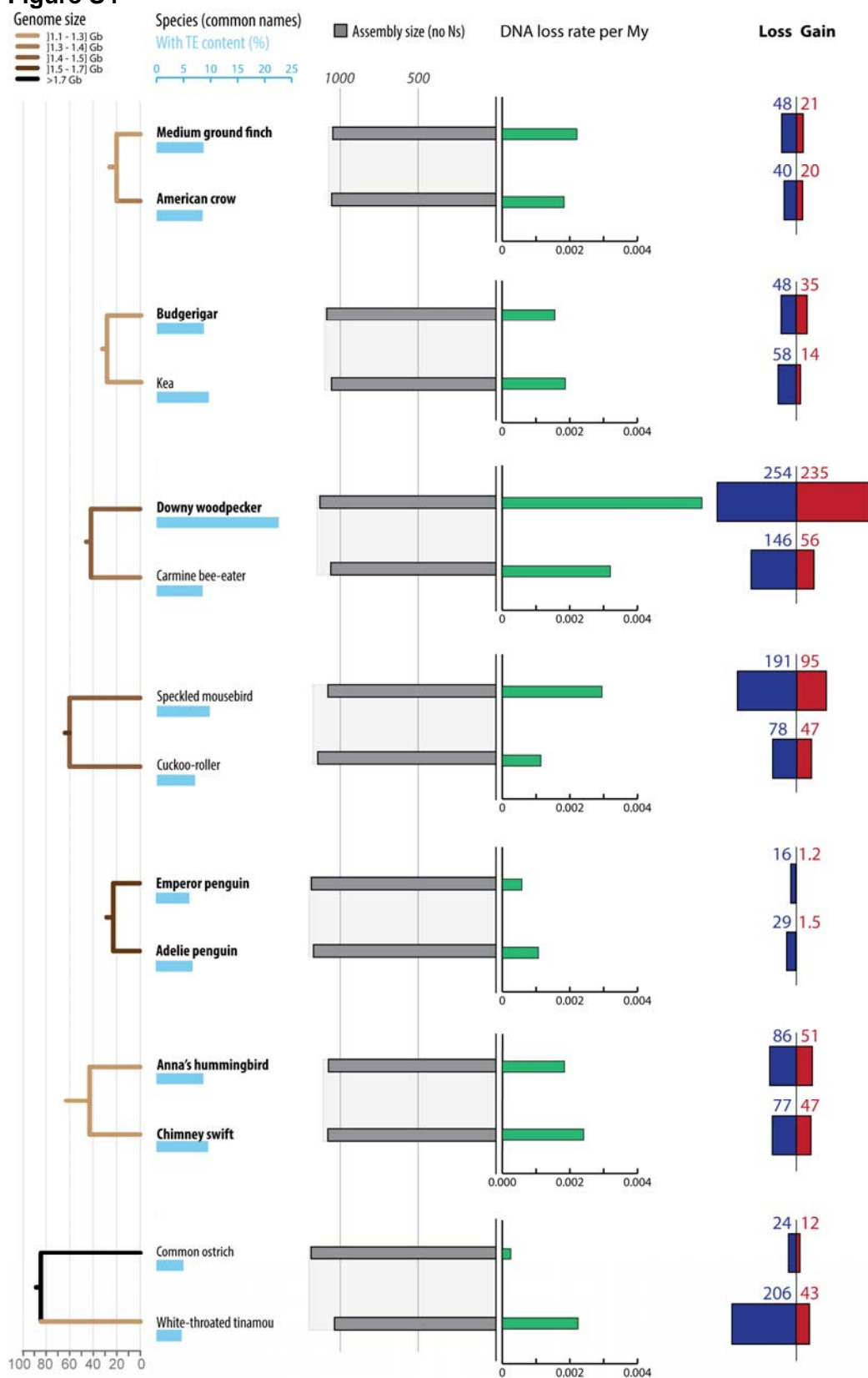
Figure S4: Gain and loss of DNA for seven duos of birds

Different species or different time scales than Figures 2 or S1 are considered, to match the times of microdeletion and midsize deletion rates of Figure 6. For each duo, we calculated the amounts of gain and loss of DNA after the two birds diverged. DNA gains (in red) correspond to lineage-specific TEs since the split of the last common ancestor of the three birds considered. Ancestral assembly size used for calculation of DNA loss (Methods) for each duo is shown with light grey highlight behind assembly sizes. It is based on the common ancestor genome size, which was estimated using the genome sizes of all species (combination of data from [114] and extrapolations from assembly sizes and coverage, color coded on the branches) and based on parsimony [52]. Estimations of the ancestral assembly size could be flawed and affect DNA loss estimations; but we observe that microdeletion and midsize deletion rates still mildly correlate with the DNA loss coefficients when calculated at the same time scale ($R^2 = 0.52$ with $p = 0.027$ and $R^2 = 0.44$ with $p = 0.087$ respectively; $R^2 = 0.51$ with $p = 0.046$ for both rates combined). Importantly, DNA gain and loss are still significantly correlated ($R^2 = 0.82$ with $p = 0.0003$, Dataset S1), which is indicating of a continuous counteraction of DNA gain by DNA loss. For each species, phylogenetic relationship (left panel) [49], TE content (light blue bars), assembly sizes (with N removed, grey bars), DNA loss rates per My (green bars) as well as gain (red) and loss (dark blue) of DNA are shown. Species names in bold correspond to high-coverage genomes, and the others to low-coverage genomes [40]. See Dataset S1 for numbers, calculation steps and details. All numbers are in Mb.

Indel ecd graph from MultiZ 100 species
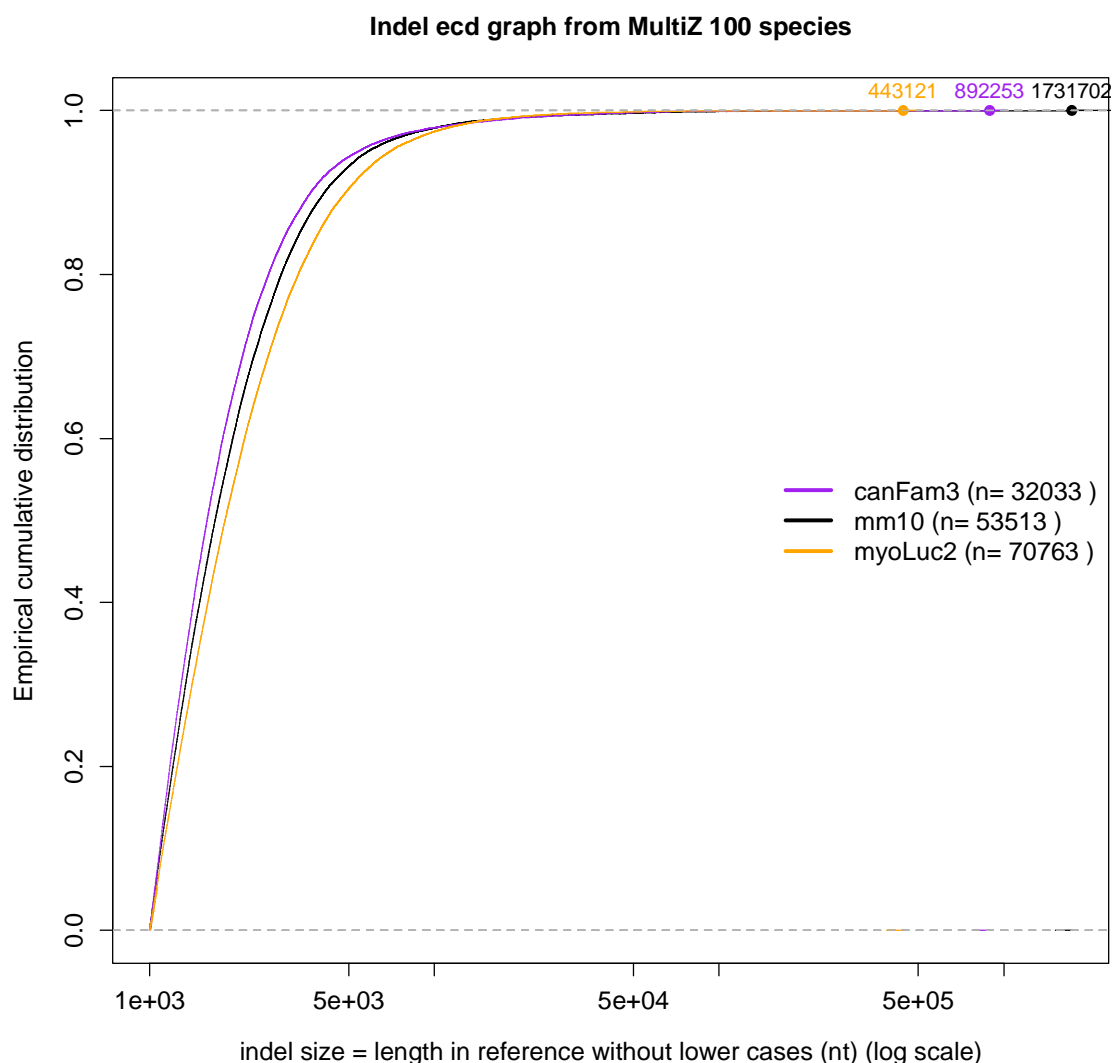
Figure S5: Size distribution of large indels in mammals

Large indels screened from the MultiZ 100-species alignment from UCSC genome browser ("empty data", see Methods). The empirical cumulative distributions are done in R [117], as in Figure S3. Statistical tests are done as in Figure S3 and are detailed in Dataset S4. Largest events are labeled on the graphs, and we verified that the three corresponding sequences in hg19 could not be found in the query species (blastn against whole-genome shotgun contigs, with repeats masked). For example, the largest deletion of the microbat is 443,121 bp long in hg19 (976,711 bp with lowercases) and is at position GL430240:588646 in myoLuc2, corresponding to chr1:238471164-239447875 in hg19. This event seems shared with other *Myotis* species and with the big brown bat *Eptesicus fuscus*, but not with the two megabats.

Table S1: List of ancient mammalian TEs

Repeat Name	Repeat Class	Repeat Family	Repeat Age (Methods)
MER67C	LTR	ERV1	Boreotheria/Eutheria?
MER67D	LTR	ERV1	Boreotheria/Eutheria?
MER70C	LTR	ERV3	Boreotheria/Eutheria?
7SK	nonTE	Pseudogene	Eutheria
7SLRNA	nonTE	srpRNA	Eutheria
Arthur1	DNA	hAT-Tip100	Eutheria
Arthur1A	DNA	hAT-Tip100	Eutheria
Arthur1B	DNA	hAT-Tip100	Eutheria
Arthur1C	DNA	hAT-Tip100	Eutheria
BLACKJACK	DNA	hAT-Blackjack	Eutheria
Chap1_Mam	DNA	hAT-Charlie	Eutheria
Charlie10	DNA	hAT-Charlie	Eutheria
Charlie10a	DNA	hAT-Charlie	Eutheria
Charlie10b	DNA	hAT-Charlie	Eutheria
Charlie11	DNA	hAT-Charlie	Eutheria
Charlie14a	DNA	hAT-Charlie	Eutheria
Charlie15a	DNA	hAT-Charlie	Eutheria
Charlie15b	DNA	hAT-Charlie	Eutheria
Charlie16a	DNA	hAT-Charlie	Eutheria
Charlie17	DNA	hAT-Charlie	Eutheria
Charlie17a	DNA	hAT-Charlie	Eutheria
Charlie17b	DNA	hAT-Charlie	Eutheria
Charlie19a	DNA	hAT-Charlie	Eutheria
Charlie20a	DNA	hAT-Charlie	Eutheria
Charlie21a	DNA	hAT-Charlie	Eutheria
Charlie22a	DNA	hAT-Charlie	Eutheria
Charlie23a	DNA	hAT-Charlie	Eutheria
Charlie24	DNA	hAT-Charlie	Eutheria
Charlie25	DNA	hAT-Charlie	Eutheria
Charlie26a	DNA	hAT-Charlie	Eutheria
Charlie29a	DNA	hAT-Charlie	Eutheria
Charlie2a	DNA	hAT-Charlie	Eutheria
Charlie2b	DNA	hAT-Charlie	Eutheria
Charlie4a	DNA	hAT-Charlie	Eutheria
Charlie4z	DNA	hAT-Charlie	Eutheria
Charlie5	DNA	hAT-Charlie	Eutheria
Charlie6	DNA	hAT-Charlie	Eutheria
Charlie7a	DNA	hAT-Charlie	Eutheria
Charlie8	DNA	hAT-Charlie	Eutheria
Charlie9	DNA	hAT-Charlie	Eutheria
Cheshire	DNA	hAT-Charlie	Eutheria
Cheshire_Mars_	DNA	hAT-Charlie	Eutheria
DNA1_Mam	DNA	TcMar	Eutheria
ERV3-16A3_I-int	LTR	ERV3/L	Eutheria
ERV3-16A3_LTR	LTR	ERV3/L	Eutheria
ERVL-B4-int	LTR	ERVL	Eutheria
ERVL-E-int	LTR	ERVL	Eutheria
ERVL-int	LTR	ERVL	Eutheria
FordPrefect	DNA	hAT-Tip100	Eutheria
FordPrefect_a	DNA	hAT-Tip100	Eutheria
HAL1	LINE	L1	Eutheria
HAL1b	LINE	L1	Eutheria
HAL1M8	LINE	L1	Eutheria
HAL1ME	LINE	L1	Eutheria
hAT-N1_Mam	DNA	hAT	Eutheria
Helitron1Na_Mam	DNA	Helitron	Eutheria
Helitron1Nb_Mam	DNA	Helitron	Eutheria
Helitron3Na_Mam	DNA	Helitron	Eutheria
HERV16-int	LTR	ERVL	Eutheria
HERVL40-int	LTR	ERVL	Eutheria
Kanga1	DNA	TcMar-Tc2	Eutheria
Kanga11a	DNA	TcMar-Tc2	Eutheria
Kanga1a	DNA	TcMar-Tc2	Eutheria

Kanga1b	DNA	TcMar-Tc2	Eutheria
Kanga1c	DNA	TcMar-Tc2	Eutheria
Kanga1d	DNA	TcMar-Tc2	Eutheria
Kanga2_a	DNA	TcMar-Tc2	Eutheria
L1M2	LINE	L1	Eutheria
L1M4	LINE	L1	Eutheria
L1M4a1	LINE	L1	Eutheria
L1M4a2	LINE	L1	Eutheria
L1M4b	LINE	L1	Eutheria
L1M4c	LINE	L1	Eutheria
L1M5	LINE	L1	Eutheria
L1M6	LINE	L1	Eutheria
L1M6B	LINE	L1	Eutheria
L1M7	LINE	L1	Eutheria
L1M8	LINE	L1	Eutheria
L1MB3	LINE	L1	Eutheria
L1MB4	LINE	L1	Eutheria
L1MB5	LINE	L1	Eutheria
L1MB7	LINE	L1	Eutheria
L1MB8	LINE	L1	Eutheria
L1MC	LINE	L1	Eutheria
L1MC1	LINE	L1	Eutheria
L1MC2	LINE	L1	Eutheria
L1MC3	LINE	L1	Eutheria
L1MC4	LINE	L1	Eutheria
L1MC4a	LINE	L1	Eutheria
L1MC5	LINE	L1	Eutheria
L1MC5a	LINE	L1	Eutheria
L1MCa	LINE	L1	Eutheria
L1MCb	LINE	L1	Eutheria
L1MCc	LINE	L1	Eutheria
L1MD	LINE	L1	Eutheria
L1MD1	LINE	L1	Eutheria
L1MD2	LINE	L1	Eutheria
L1MD3	LINE	L1	Eutheria
L1MDa	LINE	L1	Eutheria
L1MDb	LINE	L1	Eutheria
L1ME1	LINE	L1	Eutheria
L1ME2	LINE	L1	Eutheria
L1ME2z	LINE	L1	Eutheria
L1ME3	LINE	L1	Eutheria
L1ME3A	LINE	L1	Eutheria
L1ME3B	LINE	L1	Eutheria
L1ME3C	LINE	L1	Eutheria
L1ME3Cz	LINE	L1	Eutheria
L1ME3D	LINE	L1	Eutheria
L1ME3E	LINE	L1	Eutheria
L1ME3F	LINE	L1	Eutheria
L1ME3G	LINE	L1	Eutheria
L1ME4a	LINE	L1	Eutheria
L1ME4b	LINE	L1	Eutheria
L1ME4c	LINE	L1	Eutheria
L1ME5	LINE	L1	Eutheria
L1MEa	LINE	L1	Eutheria
L1MEb	LINE	L1	Eutheria
L1MEc	LINE	L1	Eutheria
L1MEd	LINE	L1	Eutheria
L1MEf	LINE	L1	Eutheria
L1MEg	LINE	L1	Eutheria
L1MEg1	LINE	L1	Eutheria
L1MEg2	LINE	L1	Eutheria
L1MEh	LINE	L1	Eutheria
L1MEi	LINE	L1	Eutheria
L1PB4	LINE	L1	Eutheria
L4_A_Mam	LINE	RTEX	Eutheria
L4_B_Mam	LINE	RTEX	Eutheria

L4_C_Mam	LINE	RTEX	Eutheria
Looper	DNA	PiggyBac	Eutheria
LTR101_Mam	LTR	ERVL	Eutheria
LTR102_Mam	LTR	ERVL	Eutheria
LTR103_Mam	LTR	ERV1?	Eutheria
LTR103b_Mam	LTR	ERV1?	Eutheria
LTR104_Mam	LTR	Gypsy	Eutheria
LTR105_Mam	LTR	ERVL	Eutheria
LTRMam	LTR	LTR	Eutheria
LTR107_Mam	LTR	LTR	Eutheria
LTR108a_Mam	LTR	ERVL	Eutheria
LTR108b_Mam	LTR	ERVL	Eutheria
LTR108c_Mam	LTR	ERVL	Eutheria
LTR108d_Mam	LTR	ERVL	Eutheria
LTR108e_Mam	LTR	ERVL	Eutheria
LTR16	LTR	ERV3	Eutheria
LTR16A	LTR	ERV3	Eutheria
LTR16A1	LTR	ERV3	Eutheria
LTR16A2	LTR	ERV3	Eutheria
LTR16B	LTR	ERV3	Eutheria
LTR16B1	LTR	ERV3	Eutheria
LTR16B2	LTR	ERV3	Eutheria
LTR16C	LTR	ERV3	Eutheria
LTR16D	LTR	ERV3	Eutheria
LTR16D1	LTR	ERV3	Eutheria
LTR16D2	LTR	ERV3	Eutheria
LTR16E1	LTR	ERV3	Eutheria
LTR16E2	LTR	ERV3	Eutheria
LTR33	LTR	ERV3	Eutheria
LTR33A	LTR	ERV3	Eutheria
LTR33A_	LTR	ERVL	Eutheria
LTR33B	LTR	ERV3	Eutheria
LTR33C	LTR	ERV3	Eutheria
LTR37A	LTR	ERV1	Eutheria
LTR37B	LTR	ERV1	Eutheria
LTR37-int	LTR	ERV1	Eutheria
LTR40a	LTR	ERV3	Eutheria
LTR40A1	LTR	ERV3	Eutheria
LTR40b	LTR	ERV3	Eutheria
LTR40c	LTR	ERV3	Eutheria
LTR41	LTR	ERV3	Eutheria
LTR41B	LTR	ERV3	Eutheria
LTR41C	LTR	ERV3	Eutheria
LTR45	LTR	ERV1	Eutheria
LTR50	LTR	ERV3	Eutheria
LTR52	LTR	ERV3	Eutheria
LTR52-int	LTR	ERVL	Eutheria
LTR53	LTR	ERV3	Eutheria
LTR53B	LTR	ERV3	Eutheria
LTR55	LTR	ERV3	Eutheria
LTR58	LTR	ERV1	Eutheria
LTR65	LTR	ERV1	Eutheria
LTR67B	LTR	ERV3	Eutheria
LTR68	LTR	ERV1	Eutheria
LTR69	LTR	ERV3	Eutheria
LTR73	LTR	ERV1	Eutheria
LTR75	LTR	ERV3	Eutheria
LTR75_1	LTR	ERV1	Eutheria
LTR75B	LTR	ERV3	Eutheria
LTR77	LTR	ERV1	Eutheria
LTR78	LTR	ERV1	Eutheria
LTR78B	LTR	ERV1	Eutheria
LTR79	LTR	ERV3	Eutheria
LTR80A	LTR	ERV3	Eutheria
LTR80B	LTR	ERV3	Eutheria
LTR81	LTR	Gypsy	Eutheria

LTR81A	LTR	Gypsy	Eutheria
LTR81AB	LTR	ERV1	Eutheria
LTR81B	LTR	ERV1	Eutheria
LTR81C	LTR	ERV1	Eutheria
LTR82A	LTR	ERV3	Eutheria
LTR82B	LTR	ERV3	Eutheria
LTR83	LTR	ERV3	Eutheria
LTR84a	LTR	ERV3	Eutheria
LTR84b	LTR	ERV3	Eutheria
LTR85a	LTR	Gypsy?	Eutheria
LTR85b	LTR	Gypsy?	Eutheria
LTR85c	LTR	Gypsy?	Eutheria
LTR86A1	LTR	ERV3	Eutheria
LTR86A2	LTR	ERV3	Eutheria
LTR86B1	LTR	ERV3	Eutheria
LTR86B2	LTR	ERV3	Eutheria
LTR86C	LTR	ERV3	Eutheria
LTR87	LTR	ERV3	Eutheria
LTR88a	LTR	Gypsy?	Eutheria
LTR88b	LTR	Gypsy?	Eutheria
LTR88c	LTR	Gypsy?	Eutheria
LTR89	LTR	ERV3	Eutheria
LTR91	LTR	LTR	Eutheria
MADE2	DNA	TcMar-Mariner	Eutheria
MamGypLTR1a	LTR	Gypsy	Eutheria
MamGypLTR1b	LTR	Gypsy	Eutheria
MamGypLTR1c	LTR	Gypsy	Eutheria
MamGypLTR1d	LTR	Gypsy	Eutheria
MamGypLTR2b	LTR	Gypsy	Eutheria
MamGypLTR2c	LTR	Gypsy	Eutheria
MamGypLTR3	LTR	Gypsy	Eutheria
MamGypLTR3a	LTR	Gypsy	Eutheria
MamRep1151	LTR?	LTR?	Eutheria
MamRep1527	Other	LTR?	Eutheria
MamRep1879	DNA	hAT?	Eutheria
MamRep1879	DNA	hAT?	Eutheria
MamRep1894	DNA	hAT	Eutheria
MamRep38	DNA	hAT	Eutheria
MamRep4096	DNA	hAT?	Eutheria
MamRep4096	DNA	hAT?	Eutheria
MamRep488	DNA	hAT	Eutheria
MamRep488	DNA	hAT	Eutheria
MamRep564	Other	ConsRep	Eutheria
MamRep605	Other	ConsRep	Eutheria
MamTip1	DNA	hAT-Tip100	Eutheria
MamTip3	DNA	hAT-Tip100	Eutheria
MARNA	DNA	TcMar-Mariner	Eutheria
MER102a	DNA	hAT-Charlie	Eutheria
MER102b	DNA	hAT-Charlie	Eutheria
MER102c	DNA	hAT-Charlie	Eutheria
MER103C	DNA	hAT-Charlie	Eutheria
MER104	DNA	TcMar-Tc2	Eutheria
MER105	DNA	DNA	Eutheria
MER105	DNA	DNA	Eutheria
MER106A	DNA	hAT-Charlie	Eutheria
MER106B	DNA	hAT-Charlie	Eutheria
MER110	LTR	ERV1	Eutheria
MER110A	LTR	ERV1	Eutheria
MER110-int	LTR	ERV1	Eutheria
MER112	DNA	hAT-Charlie	Eutheria
MER113	DNA	hAT-Charlie	Eutheria
MER113A	DNA	hAT-Charlie	Eutheria
MER113B	DNA	hAT-Charlie	Eutheria
MER115	DNA	hAT-Tip100	Eutheria
MER117	DNA	hAT-Charlie	Eutheria
MER119	DNA	hAT-Charlie	Eutheria

MER20	DNA	hAT-Charlie	Eutheria
MER20B	DNA	hAT-Charlie	Eutheria
MER21B	LTR	ERV3	Eutheria
MER21C	LTR	ERV3	Eutheria
MER21-int	LTR	ERVL	Eutheria
MER2B	DNA	TcMar-Tigger	Eutheria
MER3	DNA	hAT-Charlie	Eutheria
MER31-int	LTR	ERV1	Eutheria
MER33	DNA	hAT-Charlie	Eutheria
MER34A	LTR	ERV1	Eutheria
MER34A1	LTR	ERV1	Eutheria
MER45A	DNA	hAT-Tip100	Eutheria
MER45B	DNA	hAT-Tip100	Eutheria
MER45C	DNA	hAT-Tip100	Eutheria
MER45R	DNA	hAT-Tip100	Eutheria
MER46C	DNA	TcMar-Tigger	Eutheria
MER4B-int	LTR	ERV1	Eutheria
MER52-int	LTR	ERV1	Eutheria
MER53	DNA	hAT	Eutheria
MER58A	DNA	hAT-Charlie	Eutheria
MER58B	DNA	hAT-Charlie	Eutheria
MER58C	DNA	hAT-Charlie	Eutheria
MER58D	DNA	hAT-Charlie	Eutheria
MER5A	DNA	hAT-Charlie	Eutheria
MER5A1	DNA	hAT-Charlie	Eutheria
MER5B	DNA	hAT-Charlie	Eutheria
MER5C	DNA	hAT-Charlie	Eutheria
MER5C1	DNA	hAT-Charlie	Eutheria
MER63A	DNA	hAT-Blackjack	Eutheria
MER63B	DNA	hAT-Blackjack	Eutheria
MER63C	DNA	hAT-Blackjack	Eutheria
MER63D	DNA	hAT-Blackjack	Eutheria
MER68	LTR	ERVL	Eutheria
MER68B	LTR	ERV3	Eutheria
MER68C	LTR	ERV3	Eutheria
MER68-int	LTR	ERVL	Eutheria
MER70-int	LTR	ERVL	Eutheria
MER76	LTR	ERV3	Eutheria
MER76-int	LTR	ERV3	Eutheria
MER77	LTR	ERV3	Eutheria
MER77B	LTR	ERVL	Eutheria
MER81	DNA	hAT-Blackjack	Eutheria
MER83A-int	LTR	ERV1	Eutheria
MER83B-int	LTR	ERV1	Eutheria
MER88	LTR	ERV3	Eutheria
MER89	LTR	ERV1	Eutheria
MER89-int	LTR	ERV1	Eutheria
MER90	LTR	ERV1	Eutheria
MER90a	LTR	ERV1	Eutheria
MER91A	DNA	hAT-Tip100	Eutheria
MER91B	DNA	hAT-Tip100	Eutheria
MER91C	DNA	hAT-Tip100	Eutheria
MER92A	LTR	ERV1	Eutheria
MER92B	LTR	ERV1	Eutheria
MER92C	LTR	ERV1	Eutheria
MER92D	LTR	ERV1	Eutheria
MER92-int	LTR	ERV1	Eutheria
MER94	DNA	hAT-Blackjack	Eutheria
MER94B	DNA	hAT-Blackjack	Eutheria
MER96	DNA	hAT	Eutheria
MER96	DNA	hAT	Eutheria
MER96B	DNA	hAT	Eutheria
MER96B	DNA	hAT	Eutheria
MER97a	DNA	hAT-Tip100	Eutheria
MER97b	DNA	hAT-Tip100	Eutheria
MER97c	DNA	hAT-Tip100	Eutheria

MER97d	DNA	hAT-Tip100	Eutheria
MER99	DNA	hAT?	Eutheria
MERX	DNA	TcMar-Tigger	Eutheria
MLT1B	LTR	ERV3	Eutheria
MLT1B-int	LTR	ERVL-MaLR	Eutheria
MLT1C	LTR	ERV3	Eutheria
MLT1C-int	LTR	ERVL-MaLR	Eutheria
MLT1D	LTR	ERV3	Eutheria
MLT1D-int	LTR	ERVL-MaLR	Eutheria
MLT1E	LTR	ERV3	Eutheria
MLT1E1	LTR	ERV3	Eutheria
MLT1E1A	LTR	ERV3	Eutheria
MLT1E1A-int	LTR	ERVL-MaLR	Eutheria
MLT1E1-int	LTR	ERVL-MaLR	Eutheria
MLT1E2	LTR	ERV3	Eutheria
MLT1E2-int	LTR	ERVL-MaLR	Eutheria
MLT1E3	LTR	ERVL-MaLR	Eutheria
MLT1E3-int	LTR	ERVL-MaLR	Eutheria
MLT1E-int	LTR	ERVL-MaLR	Eutheria
MLT1F	LTR	ERV3	Eutheria
MLT1F1	LTR	ERV3	Eutheria
MLT1F1-int	LTR	ERVL-MaLR	Eutheria
MLT1F2	LTR	ERV3	Eutheria
MLT1F2-int	LTR	ERVL-MaLR	Eutheria
MLT1F-int	LTR	ERVL-MaLR	Eutheria
MLT1G	LTR	ERV3	Eutheria
MLT1G1	LTR	ERV3	Eutheria
MLT1G1-int	LTR	ERVL-MaLR	Eutheria
MLT1G3	LTR	ERV3	Eutheria
MLT1G3-int	LTR	ERVL-MaLR	Eutheria
MLT1G-int	LTR	ERVL-MaLR	Eutheria
MLT1H	LTR	ERV3	Eutheria
MLT1H1	LTR	ERV3	Eutheria
MLT1H1-int	LTR	ERVL-MaLR	Eutheria
MLT1H2	LTR	ERV3	Eutheria
MLT1H2-int	LTR	ERVL-MaLR	Eutheria
MLT1H-int	LTR	ERVL-MaLR	Eutheria
MLT1I	LTR	ERV3	Eutheria
MLT1I-int	LTR	ERVL-MaLR	Eutheria
MLT1-int	LTR	ERVL-MaLR	Eutheria
MLT1J	LTR	ERV3	Eutheria
MLT1J1	LTR	ERV3	Eutheria
MLT1J1-int	LTR	ERVL-MaLR	Eutheria
MLT1J2	LTR	ERV3	Eutheria
MLT1J2-int	LTR	ERVL-MaLR	Eutheria
MLT1J-int	LTR	ERV3	Eutheria
MLT1K	LTR	ERV3	Eutheria
MLT1K-int	LTR	ERVL-MaLR	Eutheria
MLT1L	LTR	ERV3	Eutheria
MLT1L-int	LTR	ERVL-MaLR	Eutheria
MLT1M	LTR	ERV3	Eutheria
MLT1M-int	LTR	ERVL-MaLR	Eutheria
MLT1N2	LTR	ERV3	Eutheria
MLT1N2-int	LTR	ERVL-MaLR	Eutheria
MLT1O	LTR	ERVL-MaLR	Eutheria
MLT1O-int	LTR	ERVL	Eutheria
MLT2C1	LTR	ERVL	Eutheria
MLT2C2	LTR	ERV3	Eutheria
MLT2D	LTR	ERV3	Eutheria
MLT2E	LTR	ERV3	Eutheria
MLT2F	LTR	ERV3	Eutheria
MLT-int	LTR	ERV3	Eutheria
ORSL	DNA	hAT-Tip100	Eutheria
ORSL-2a	DNA	hAT-Tip100	Eutheria
ORSL-2b	DNA	hAT-Tip100	Eutheria
Tigger11a	DNA	TcMar-Tigger	Eutheria

Tigger12	DNA	TcMar-Tigger	Eutheria
Tigger12A	DNA	TcMar-Tigger	Eutheria
Tigger12c	DNA	TcMar-Tigger	Eutheria
Tigger13a	DNA	TcMar-Tigger	Eutheria
Tigger17a	DNA	TcMar-tigger	Eutheria
Tigger17b	DNA	TcMar-tigger	Eutheria
Tigger18a	DNA	TcMar-tigger	Eutheria
Tigger20a	DNA	TcMar-tigger	Eutheria
Tigger6a	DNA	TcMar-Tigger	Eutheria
Tigger6b	DNA	TcMar-Tigger	Eutheria
Tigger8	DNA	TcMar-Tigger	Eutheria
Tigger9a	DNA	TcMar-Tigger	Eutheria
Tigger9b	DNA	TcMar-Tigger	Eutheria
Zaphod	DNA	hAT-Tip100	Eutheria
Zaphod2	DNA	hAT-Tip100	Eutheria
Zaphod3	DNA	hAT-Tip100	Eutheria
Charlie12	DNA	hAT-Charlie	Eutheria/Primates
Charlie1	DNA	hAT-Charlie	Eutheria+Metatheria
Charlie18a	DNA	hAT-Charlie	Eutheria+Metatheria
Charlie4	DNA	hAT-Charlie	Eutheria+Metatheria
Charlie7	DNA	hAT-Charlie	Eutheria+Metatheria
L2	LINE	L2	Eutheria+Metatheria
L2c	LINE	L2	Eutheria+Metatheria
Tigger17	DNA	TcMar-tigger	Eutheria+Metatheria
MamRep137	DNA	TcMar?	Eutheria/Theria
MamRep137	DNA	TcMar?	Eutheria/Theria
L1M	LINE	L1	Eutheria/Theria?
L2a	LINE	L2	Eutheria/Theria?
L2b	LINE	L2-CR1	Eutheria/Theria?
Charlie1a	DNA	hAT-Charlie	Eutheria
Charlie1b	DNA	hAT-Charlie	Eutheria
Charlie1b_Mars	DNA	hAT-Charlie	Eutheria
LTR90A	LTR	LTR	Eutheria
LTR90B	LTR	LTR	Eutheria
X7A_LINE	LINE	CR1	Amniota/Theria
X7B_LINE	LINE	CR1	Amniota/Theria
X7C_LINE	LINE	CR1	Amniota/Theria
X7D_LINE	LINE	CR1	Amniota/Theria
Helitron2Na_Mam	DNA	Helitron?	Theria/Mammalia
L1MEj	LINE	L1	Theria/Mammalia
L3	LINE	CR1	Theria/Mammalia
L3b	LINE	CR1	Theria/Mammalia
L5	LINE	RTEX	Theria/Mammalia
MamRep434	DNA	TcMar-Tigger	Theria/Mammalia
MamTip2	DNA	hAT-Tip100	Theria/Mammalia
MER124	DNA	DNA	Theria/Mammalia
Tigger10	DNA	TcMar-Tigger	Theria/Mammalia
Tigger14a	DNA	TcMar-Tigger	Theria/Mammalia
Tigger15a	DNA	TcMar-Tigger	Theria/Mammalia
Tigger16a	DNA	TcMar-Tigger	Theria/Mammalia
Tigger16b	DNA	TcMar-Tigger	Theria/Mammalia
Tigger19a	DNA	TcMar-tigger	Theria/Mammalia
Charlie13a	DNA	hAT-Charlie	Mammalia
Charlie13b	DNA	hAT-Charlie	Mammalia
CR1_Mam	LINE	CR1	Mammalia
Mam_R4	LINE	Dong-R4	Mammalia
MamGyp-int	LTR	Gypsy	Mammalia
MamRep1161	DNA	TcMar	Mammalia
MamSINE1	SINE	tRNA	Mammalia
MER121	DNA	TcMar? hAT?	Mammalia
MER135	DNA	DNA	Mammalia
MIR	SINE	MIR	Mammalia
MIR3	SINE	MIR	Mammalia
MIRb	SINE	MIR	Mammalia
MIRc	SINE	MIR	Mammalia
Plat_L3	LINE	CR1	Mammalia

X3_LINE	LINE	RTE-BovB	Mammalia
X6A_LINE	LINE	CR1	Mammalia
X6B_LINE	LINE	CR1	Mammalia
AmnSINE1	SINE	Deu	Amniota
AmnSINE2	SINE	Deu	Amniota
Eulor1	DNA	ConsRep	Amniota
Eulor10	Other	DNA?	Amniota
Eulor11	DNA	ConsRep	Amniota
Eulor12	Other	DNA?	Amniota
Eulor2A	Other	DNA?	Amniota
Eulor2B	Other	DNA?	Amniota
Eulor2C	Other	DNA?	Amniota
Eulor3	Other	DNA?	Amniota
Eulor7	Other	DNA?	Amniota
Eulor8	DNA	TcMar?	Amniota
Eulor9A	DNA	ConsRep	Amniota
Eulor9B	DNA	ConsRep	Amniota
Eulor9C	DNA	ConsRep	Amniota
MER125	DNA	DNA	Amniota
MER126	DNA	DNA	Amniota
MER127	DNA	TcMar-Tigger	Amniota
MER129	SINE	SINE	Amniota
MER130	Other	ConsRep	Amniota
MER131	SINE?	SINE?	Amniota
MER132	DNA	TcMar	Amniota
MER132	DNA	TcMar	Amniota
MER133A	DNA	DNA?	Amniota
MER133B	DNA	DNA?	Amniota
MER134	DNA	DNA	Amniota
MER136	DNA	DNA	Amniota
UCON12	Other	ConsRep	Amniota
UCON12A	Other	ConsRep	Amniota
UCON15	Other	ConsRep	Amniota
UCON16	Other	ConsRep	Amniota
UCON17	Other	ConsRep	Amniota
UCON19	Other	ConsRep	Amniota
UCON2	Other	ConsRep	Amniota
UCON20	Other	ConsRep	Amniota
UCON21	Other	DNA?	Amniota
UCON22	Other	ConsRep	Amniota
UCON23	DNA	hAT-Tip100?	Amniota
UCON24	Other	ConsRep	Amniota
UCON25	Other	ConsRep	Amniota
UCON26	Other	ConsRep	Amniota
UCON27	Other	ConsRep	Amniota
UCON28a	Other	ConsRep	Amniota
UCON28b	Other	ConsRep	Amniota
UCON28c	Other	ConsRep	Amniota
UCON29	DNA	PiggyBac?	Amniota
UCON4	Other	ConsRep	Amniota
UCON5	Other	ConsRep	Amniota
UCON6	Other	ConsRep	Amniota
UCON9	Other	ConsRep	Amniota
X1_LINE	LINE	CR1	Amniota
X8_LINE	LINE	CR1	Amniota
X9_LINE	LINE	CR1	Amniota
X2_LINE	LINE	CR1	Amniota/Euteleostemi?
Eulor4	Other	ConsRep	Tetrapoda
Eulor5A	Other	DNA?	Tetrapoda
Eulor5B	Other	DNA?	Tetrapoda
Eulor6A	Other	DNA?	Tetrapoda
Eulor6B	Other	DNA?	Tetrapoda
Eulor6C	Other	DNA?	Tetrapoda
Eulor6D	Other	DNA?	Tetrapoda
Eulor6E	Other	DNA?	Tetrapoda
LFSINE_Vert	SINE	SINE	Tetrapoda

MER123	DNA	DNA	Tetrapoda
UCON1	Other	ConsRep	Tetrapoda
UCON10	Other	ConsRep	Tetrapoda
UCON11	DNA	TcMar-tigger	Tetrapoda
UCON13	Other	Penelope	Tetrapoda
UCON14	Other	DNA?	Tetrapoda
UCON18	Other	ConsRep	Tetrapoda
UCON31	Other	ConsRep	Tetrapoda
UCON7	Other	DNA?	Tetrapoda
UCON8	DNA	DNA	Tetrapoda
X5A_LINE	LINE	CR1	Euteleostomi
X5B_LINE	LINE	CR1	Euteleostomi