Dynamics of genome size evolution in birds and mammals

Aurélie Kapusta[a], Alexander Suh[b], and Cédric Feschotte[a,1]

[a]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112; and [b]Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden

Genome size in mammals and birds shows remarkably little inter-specific variation compared with other taxa. However, 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 covariation between the amount of DNA gained by transposition and lost by deletion during mammal and avian evolution, resulting in genome size equilibrium. To test this model, we develop computational methods to quantify the amount of DNA gained by TE expansion and lost by deletion over the last 100 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 extents 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.

genome size evolution | DNA loss | transposable elements | amniotes | flight

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 (1–7). 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). Additionally, the fixation probability of slightly deleterious deletions or insertions would be higher in species with smaller effective population sizes, where natural selection acts less efficiently (3, 8). On the other hand, a number of correlative associations between genome size and phenotypic traits, such as cell size (9, 10) and metabolic rate associated with powered flight (11, 12), 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., refs. 1, 2, 13, and 14).

Most studies of genome size evolution have focused on taxa with extensive variation in genome sizes, such as flowering plants (15–20), conifers (21), insects (22–25), teleost fishes (26), or species with extreme sizes [such as pufferfishes (27, 28) and salamanders (29, 30)]. 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 (for reviews, see refs. 2, 5, and 31). 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 with these taxa, birds and mammals show little interspecific variation in genome size (from ~1–2.1 Gb and 1.6–6.3 Gb, respectively) (Fig. 1), and little is known about the mechanisms underlying genome size equilibrium 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 (35). 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 (36–47). 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 equilibrium in mammals and birds (as hypothesized in refs. 14 and 48). 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 (49) implicated that variation in DNA gains through TE expansion acted in concert with variations
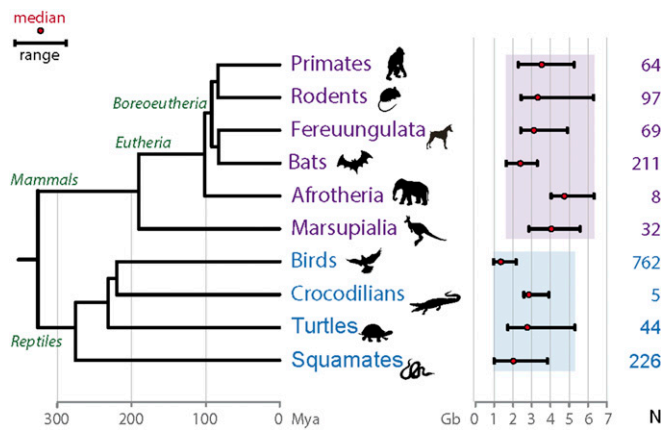
Fig. 1. Genome size variation in amniotes. Cytological haploid genome size ranges of different groups of species are shown as black bars (from smallest to largest genome sizes). Birds range from ~0.96–2.2 Gb (~2.25×), whereas all reptiles range from 1.1 to 5.44 Gb (~5×, shown as a blue rectangle) and mammals from 1.6 to 6.3 Gb (~3.9×, shown as a purple rectangle). For rodents, the red viscacha rat (tetraploid, 8.4 Gb) was not included. This is in contrast to the ranges among all vertebrates (~0.35 to ~133 Gb, ~333×, 3,731 species), or among the other classes of vertebrates that include more than one family: amphibians (0.95–121 Gb, ~127×, 504 species), cartilaginous fishes (Chondrichthyes; 1.5–17 Gb, ~11×, 134 species) and bony fishes (Osteichthyes; 0.35–133 Gb, ~379.5×, 1,407 species). We note that among the 25 orders of Osteichthyes with more than 4 species, 6 have a genome size range >5× (total of 897 species), with a maximum of ~8.7× (Cypriniformes, 229 species). The average of the 25 within-Osteichthyes genome size ranges is ~3.7×, which is a similar range than the one of mammals without the red viscacha rat (32). Divergence times are represented on the phylogenetic tree in millions of years as in refs. 33 and 34. Red dot: median. N: numbers of species inside each group with genome size data included in the figure (compiled from ref. 32, as of March 6, 2015). When several measures exist for one single species, values are averaged. Gb: gigabases.

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 (39) showed that the dog and human lineages experienced 2.5× less DNA loss than in the mouse lineage, but also 2.8× and 1.6× less DNA gain, a balance explaining the modest differences in genome size across these species. In birds, little is known about genome size dynamics. Statistical models of genome size evolution in the avian lineage have inferred that a contraction of ~0.8-fold occurred before the divergence of birds in a theropod ancestor (50). Consistent with this idea, a recent comparative analysis of 48 bird genomes revealed that introns and intergenic regions are, on average, ~2× and ~1.4× smaller in birds than in mammals and nonavian reptiles, respectively (46) (see also refs. 51 and 52). Furthermore, the ostrich lineage was found to have experienced, on average, larger genomic deletions than alligator and turtle lineages (46). 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., ref. 53) and avian (46) 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 equilibrium, 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 Dataset S1). Using this information, we applied the Repeat-Masker software (54) 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) (37, 40, 41, 43, 44). 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 (46) (*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 (55).

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) from 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 because of 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 (56, 57). 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 (58), 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 S2).

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 ref. 39) 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 (100 My for mammals and 70 My for birds) (*Methods* and Dataset S2). Applying our predicted DNA loss rate for human, we obtain a coefficient k of 0.0026 per million years, which is nearly identical to the previously calculated coefficient of 0.0024 (39), despite being based on a different methodology to infer DNA gain and loss.

When applied to all lineages (Figs. 2 and 3 and Fig. 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 sixfold across mammals (from 150 Mb in the megabat to 1,007 Mb in the mouse lineages) (Fig. 2) and by more than 30-fold across birds (from 7 Mb in the ostrich to 255 Mb in the woodpecker lineages) (Fig. 3). DNA loss amounts range by twofold across mammals (from 650 Mb in the human to 1,373 Mb in the microbat lineages) and by more than threefold across birds (from 119 Mb in the ostrich to 424 Mb in the woodpecker lineages).
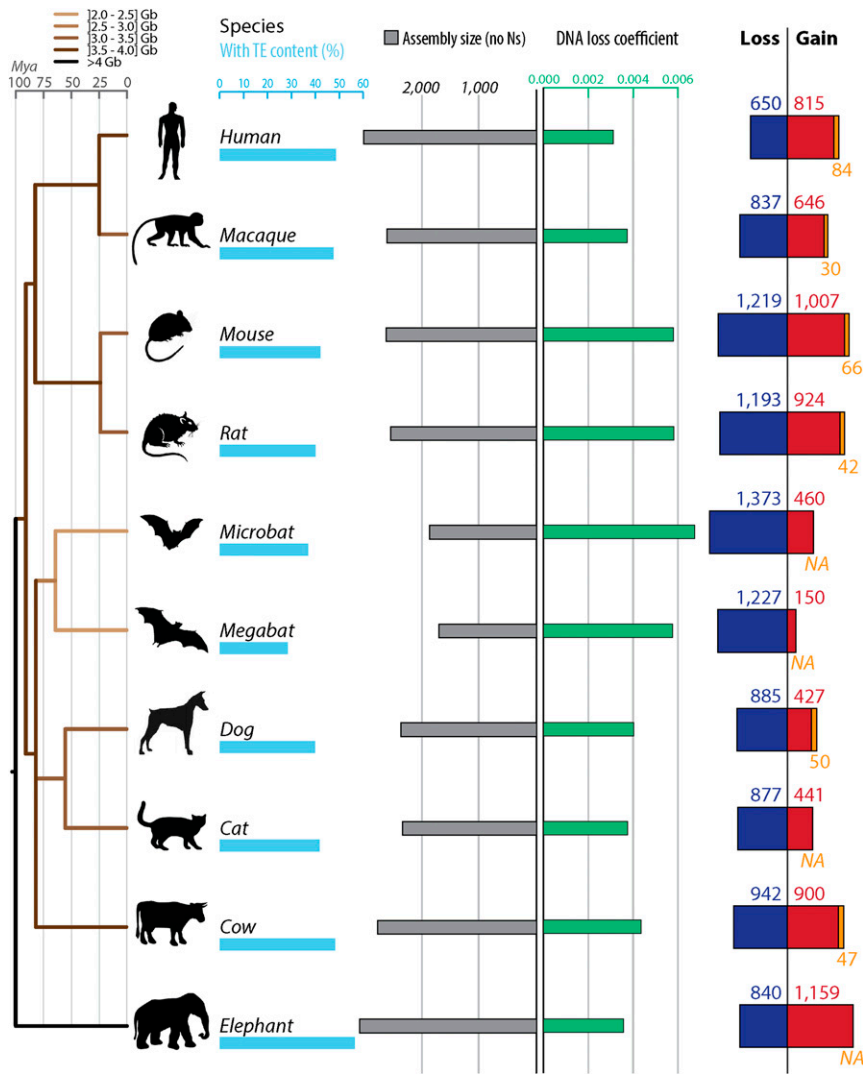
Fig. 2. Gain and loss of DNA in 10 mammalian lineages. For each species, phylogenetic relationship (*Left*) (34), TE content (light blue bars), assembly sizes (with *N* removed, gray bars), DNA loss coefficients (green bars), as well as gain (red and orange) and loss (dark blue) of DNA are shown. DNA gains correspond mostly to lineage-specific TEs in red (not shared with other mammals). When available, measures of segmental duplications were added (orange). Because segmental duplications also contain TEs, we corrected the segmental duplication amounts with the TE content of each genomes. DNA loss amounts and coefficients are calculated as in ref. 39 using a common ancestor "assembly" size for all mammals of 2,800 Mb (*Methods*). The phylogenetic tree is color-coded based on genome sizes in gigabases (data in picograms from ref. 32) based on parsimony. See Dataset S2 for numbers, calculation steps and assemblies details. All numbers are in megabases.

For mammals, these results confirm the trends previously reported for some of these lineages (36, 37, 39): 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 (Figs. 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 (Figs. 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 S2). 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) (Figs. 2 and 3 and Fig. S1).

To investigate the extent by which these two opposite forces each contribute to genome size equilibrium, we next examine whether DNA gains (percentage of the ancestor assembly size) and DNA loss coefficients correlate with assembly sizes (using Felsenstein's independent contrasts to account for phylogenetic dependence; see *Methods*). In mammals, we observe significant correlations of the contrasts in assembly sizes with the contrasts in DNA gains as well as with the contrasts in DNA loss coefficients (Pearson coefficient $r = 0.86$ with $P = 0.001$ and $r = -0.74$ with $P = 0.015$, respectively)

(Dataset S2). However, these results have to be interpreted with caution because our statistical power is limited by the relatively small sample of mammalian lineages examined ($n = 10$). For the 24 bird lineages analyzed, we observe that the contrasts in DNA loss coefficients, but not the contrasts in DNA gains, significantly correlate with the contrasts in assembly sizes (Pearson coefficient $r = -0.73$ with $P = 6.00e-05$ and $r = -0.13$ with $P = 0.54$, respectively) (Dataset S2). This observation indicates that DNA loss is a predominant contributor to genome size equilibrium in birds. However, neither one of these two forces alone can fully account for the variation between extant assembly sizes (e.g., woodpecker) (Dataset S2). Thus, these two forces must have acted in concert to modulate genome size throughout avian evolution. Consistent with this idea, we found that the contrasts in DNA gains and the contrasts in DNA loss coefficients are positively and significantly correlated with each other across the bird lineages examined (Pearson coefficient $r = 0.77$ with $P = 2.00e-05$) (Fig. 4). These data support a model where genome size equilibrium is maintained through DNA loss counteracting the gains of DNA acquired 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 (Fig. 3, Fig. S1, and Dataset S2).

Contribution of Microdeletions to DNA Loss. Having determined that DNA loss makes an important contribution to eutherian and avian

Species
With TE content (%)

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

Mya
100 80 60 40 20 0

0 5 10 15 20

Assembly size (no Ns)
1000   500

DNA loss coefficient
0.000 0.002 0.004 0.006

Loss | Gain

Species	Loss	Gain
Downy woodpecker	424	255
Carmine bee-eater	330	90
Bar-tailed trogon	315	94
Emperor penguin	150	35
Adelie penguin	170	43
Little egret	196	61
Anna's hummingbird	316	92
Chimney swift	319	99
Chuck-Will's-widow	228	68
Chicken	343	94
White-throated tinamou	300	38
Common ostrich	119	7

Neoaves
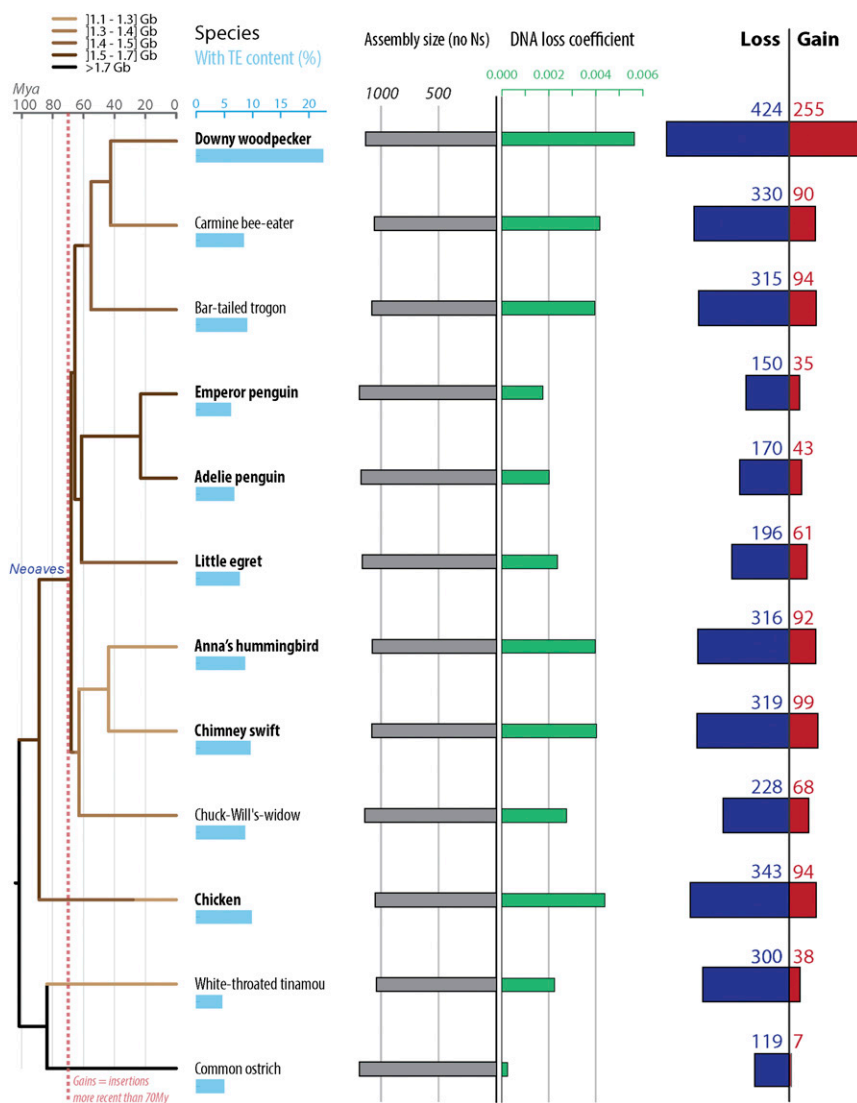
Gains = insertions more recent than 70My

Fig. 3. Gain and loss of DNA in 12 avian lineages. Twelve avian lineages mentioned in the text and with the extreme values of DNA gain and loss amounts are shown (see Fig. S1 for 12 additional species with intermediate values). For each species, phylogenetic relationship (*Left*) (55), TE content (light blue bars), assembly sizes (with *N* removed, gray 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 (46), and the others to low coverage genomes (46). DNA gain corresponds to insertions younger than 70 My. DNA loss amounts and coefficients are calculated as in ref. 39 using a common ancestor size of 1,300 Mb (see text and *Methods*). Phylogenetic tree is color coded based on genome sizes in gigabases (data in picograms from ref. 32 and extrapolations from assembly sizes and coverage), based on parsimony as in ref. 58. See Dataset S2 for numbers, calculation steps, and assemblies details. All numbers are in megabases.

genome size equilibrium, 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 University of California, Santa Cruz (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 all seven avian species included in both this alignment and in ref. 46 (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 million years (Fig. 5).

The results show that rodents have the highest microdeletion rates among the lineages examined (Fig. 5*A*), about 3.5× higher as those for the human lineage, in agreement with previous analyses of a smaller number of mammals (29, 36, 37, 62, 63). All other mammalian 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 (0.4×) microdeletion rate than that of the tenrec (*Echinops telfairi*) or that of their common ancestor. This exception 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 (for review, see ref. 64). When the number of microdeletion events (normalized by alignment length) is considered, rather than the total amount of DNA removed by microdeletions (Fig. 5*A*), 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 5.4× 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 S3). Average lengths per species lineages range from 4.8 bp (bats) to 5.7 bp (tenrec) in mammals
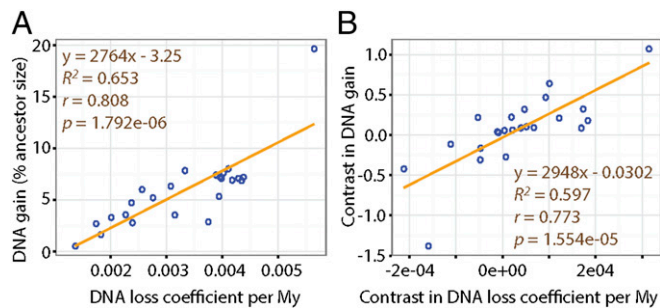
Fig. 4. Gain and loss of DNA as driving forces of genome size variation. (*A*) DNA gains (percent ancestor size) are plotted against DNA loss coefficients for the 24 birds examined. Adding coverage depth or contig N50 from ref. 46 as covariates of gain or loss does not affect the correlation (Dataset S2). (*B*) Contrasts in DNA gain (percent ancestor size) are plotted against the contrasts in DNA loss coefficients, to correct for phylogenetic relatedness (Felsenstein's independent contrasts) (see *Methods*). The lines show the linear least-squares best fits with the associated equation, squared correlation coefficient R^2, Pearson correlation coefficient r and P value. The values and R (59) command lines used to build this figure can be found in Dataset S2.

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 (Fig. S2 and Dataset S4). Within orders, pairwise species comparisons are significantly different between bats and other mammals, and between zebra finch and the other birds (Fig. S2 and Dataset S4). 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 (63). 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 (65) (*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 time frame (Figs. 2 and 3 and Fig. 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 S3).

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. 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 (Fig. 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 sequence 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 million years. To be able to compare across lineages, the deletion rates were also normalized by alignment length (shown in megabases in Fig. 6).
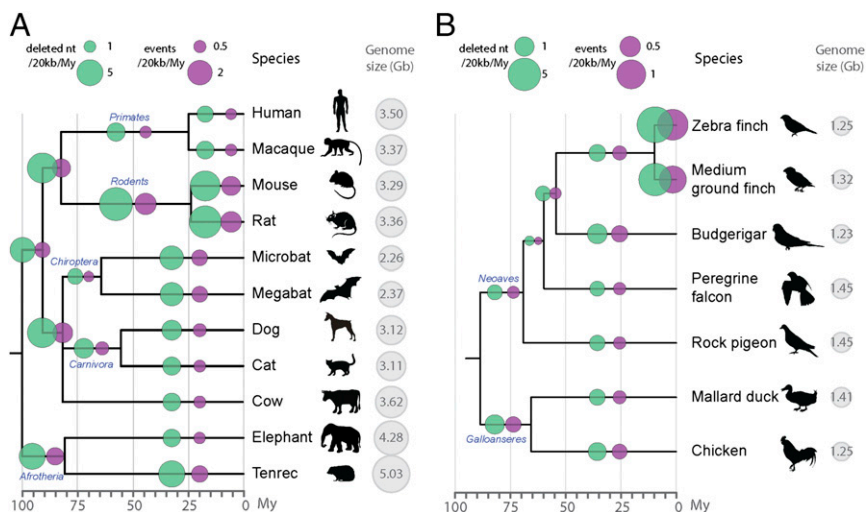

Fig. 5. Microdeletion rates across amniotes. Microdeletion (1–30 bp) rates and number of events are shown in green and purple, respectively. Microdeletions are estimated from gaps in the UCSC MultiZ 100-species alignment of human chromosomes 1–22, restricted to blocks containing information for the species studied (*Methods* and Dataset S3). Deletion rates are calculated by dividing the amount of gaps specific to each branch (not present in any other species) by millions of years of each branch. Cytological haploid genome sizes are from ref. 32. There were no reported genome sizes for two species, so the size of the closest species is shown: the Tenrecidae *Setifer setosus* for *E. telfairi* 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 *A* and *B*). (*A*) Microdeletion rates in 11 placental mammals (with *M. domestica* as outgroup). The total length of the alignment is 297 Mb and timescales are as in refs. 34, 37, 60, and 61. Names of orders are indicated on the tree (in blue). (*B*) Microdeletion rates in seven birds (with *A. carolinensis* as outgroup). The total length of the alignment is 66.5 Mb. Timescales are from ref. 55. Names of two superorders are placed on the tree (in blue).

We first used these datasets as an independent method to infer microdeletion (<30 bp) rates in nucleotides per million years 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 (Fig. 6, light green circles). For example, the elephant displays the lowest microdeletion rates in nucleotides per million years among the species compared, whereas 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× higher for mammals and 4.3× higher for birds (Dataset S5). 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 S3). 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 (Fig. 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 (Fig. 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 an ~6-kb deletion in the medium ground finch relative to the golden-collared manakin (see the legend of Fig. 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 over-estimate 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 (Fig. 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 in nucleotides per million years, respectively (Fig. 6, dark gray 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 S5) 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 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 S5). 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 in nucleotide per million years, whereas the smallest rates are observed in the ostrich and penguin lineages (Fig. 6 and Dataset S5). Next, we sought to assess the relative contribution of micro- and midsize deletions in birds to genome size equilibrium. 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 S5). We
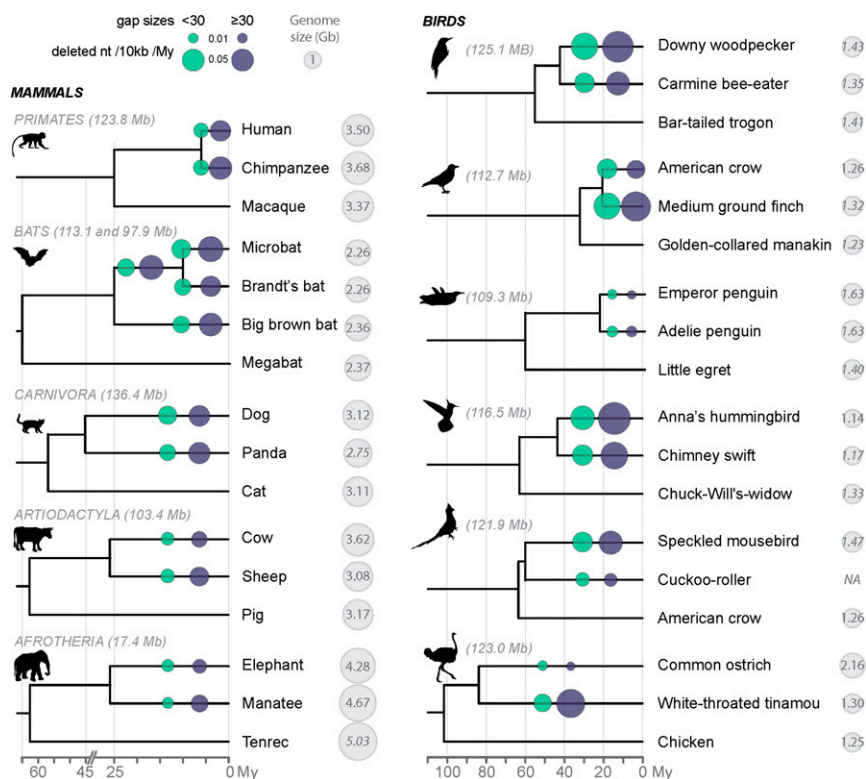

Fig. 6. Midsize deletion rates across amniotes. Microdeletions (1–30 bp) and midsize deletion (from 30 bp to 10 kb for mammals and birds, respectively) rates, measured after a recent split between two species, are shown in light green and dark blue, respectively. Deletion rates were calculated based on gaps in alignments of orthologous regions specific to the species (Methods) and are normalized by alignment length (in megabases in 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, two sets of three species were considered: group I with Myotis lucifugus, Myotis brandtii, and Eptesicus fuscus (113.1 Mb of alignment), group II with M. lucifugus, E. fuscus, and Pteropus 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 S5 for the numbers, and Fig. S3 for the midsize deletion spectrums. Cytological haploid genome sizes are from ref. 32. There were no reported genome sizes for 2 mammals and 11 bird species (genome size in italic), so the size of the closest species is shown (Dataset S5).

observe that the contrasts in the two variables are positively correlated, either if each class of deletion is considered separately ($r = 0.78$ with $P = 3.5e{-}04$ and $r = 0.73$ with $P = 0.001$ for micro- and midsize deletions, respectively) (Dataset S5) or together ($r = 0.76$ with $P = 6.2e{-}04$) (Dataset S5). These results suggest that both classes of deletions are significant drivers of genome size equilibrium 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 S5). 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 relatively little interspecific variation in genome sizes compared with others (such as plants or insects; see also legend of Fig. 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 (66–68). Our analysis shows that this is clearly not the case: there has been extensive 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, whereas the equivalent of 44% of genome content was lost over the same time frame (Fig. 2 and Dataset S2). The woodpecker lineage provides another striking example. Among birds, this species lineage has experienced the largest amount of DNA gain [255 Mb, predominantly through CR1 LINE transposition (46)] but also the largest amount of DNA loss (424 Mb, equivalent to about one-third of the genome) over the past ~70 My, resulting in a current genome size comparable to that of other modern bird species (Fig. 3, Fig. S4, and Dataset S2). 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 time frame. This accordion process helps explaining the relative maintenance of genome size across the eutherian and avian phylogeny. This is particularly evident in birds (Fig. 1), which display a positive correlation between DNA gain and DNA loss (Fig. 4). Thus, our results indicate that the relatively small genome size of birds is not merely because of a dearth of transposition in those lineages, as previously hypothesized, (e.g., refs. 66–68), but rather the result of a dynamic interplay between TE-mediated DNA acquisition and subsequent DNA loss (as suggested in refs. 14 and 48).

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 impose a relatively small upper limit for the size of observable events (because 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., refs. 22–25 and 69), plants (e.g., ref. 16), and a few vertebrates (14, 27–30). 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 refs. 70 and 71). 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., refs. 2, 26, 62, 63, and 72). 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 event can only explain a minute fraction of the DNA content lost during avian and eutherian evolution (Figs. 5 and 6 and Datasets S3 and S5) 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 (9.5–40% and 20% in average) (Dataset S5) 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 (73), and 31 kb (74); see also ref. 75], and we were able to identify individual events (Fig. 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) (46).

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 (75–77), yet it does not preclude that the process of segmental DNA loss has played an important role in driving phenotypic evolution (78–80). 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 (46, 81, 82). The foreseeable improvement of genome assembly via third-generation sequencing (e.g., long-read sequencing and gap filling; see ref. 83) will provide a way to more directly test the hypothesis that large deletion events play a prominent role in amniote genome evolution. Additionally, the resolution of tandem repeats that are generally missing from current assemblies will improve, thus enabling the quantification of their contribution to genome dynamics.

Genome Contraction Covaries with TE Expansion.

What could be the mechanisms facilitating the covariation between the amounts of DNA gained via TE insertions and the DNA that is lost, which is especially striking in bird evolution (Fig. 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, 65). Consistent with this idea, we find that microdeletion rates (and, to a lesser extent, midsize deletion rates) correlate strongly and significantly with neutral substitution rates in birds (Dataset S5). This finding 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) (8). Conversely, natural selection acts less efficiently in species with smaller effective population sizes (8), which has been suggested to contribute indirectly to the reduced purging of slightly deleterious TE insertions (3, 84; contra ref. 85). Mammals have generally smaller effective population sizes than birds (86), which is predicted to increase the probability of fixation of nearly neutral TE insertion and deletion events through genetic drift (8, 87). 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 (Figs. 2 and 3 and Fig. 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 nonallelic homologous recombination (NAHR) events that may result in the deletion of the intervening DNA (reviewed in ref. 88). Although the impact of TE-mediated NAHR on the process of DNA loss has been well-documented in plants (e.g., refs. 15, 18, 19, 89–92; or for review see ref. 31), 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 interelement NAHR deletion events to occur between TE copies. For example, the highly abundant *Alu* elements have mediated considerably more NAHR deletion events (e.g., refs. 74, 93, and 94) than L1 (95) or SVA (SINE/VNTR/Alu) (96) 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 (46) and Ves SINEs in bats (97, 98), led to an increase opportunity for NAHR, thereby facilitating the extreme degree of DNA loss that we observed in these two lineages (Figs. 2 and 3 and Fig. 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 (Figs. 2 and 3 and Figs. S1 and S4), the only vertebrates capable of powered flight. Our study also extends the previous notion that the evolution of the small genomes of bats and birds predates the emergence of flight (50, 99). The continuous genome contraction we see along multiple bird lineages (Fig. 3 and Figs. S1 and S4) is consistent with previous inference that their common ancestor had a larger genome than that of extant avian species (58). Importantly, we found no significant elevation in microdeletion rates in the respective common ancestor of bats, Paleognathae (ratites and tinamous), Galloanserae (chickens and ducks), or Neoaves (all remaining birds) (Fig. 5). Paradoxically, bats display the lowest microdeletion rate in our analysis (Fig. 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 million years) (see figure S12 of ref. 46). Together, these observations suggest that genome contraction before 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 (12, 100, 101), whereas flightless ratite birds display the largest genomes (2, 51, 102). Our results lend further support to this connection between metabolic rate and genome size reduction. We found that bird lineages that have lost flight (penguins and ostrich) are characterized by midsize deletion rates significantly lower than those of flying birds (2.3-fold on average; *ks* test, *P* = 0.0036) (Fig. 6 and Dataset S5). 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) (48). Furthermore, we observe that flightless bird lineages (penguins and ostrich) have gained generally less DNA during evolution and tend to show more older TEs than flying birds (Fig. 3, Fig. S1, and Dataset S2).

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 (103) and birds (104). 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 (Figs. 5 and 6). Similarly, megabats have larger body mass than microbats, and show a lower DNA loss coefficient (Fig. 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 S2 and S5), 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 (99, 105) (Fig. 1). We found that bats have a DNA loss/gain ratio ~4.3-fold higher than the other mammals examined (Fig. 2), as well as the highest midsize deletion rates (Fig. 6 and Dataset S5). Importantly, however, we observe that neither bats (Figs. 5A and 6) nor flying birds (Fig. 6 and Dataset S5) 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. Versions of assemblies and species names are listed in Datasets S2–S5. Genome sequences in fasta format were recovered from UCSC for mammals (hgdownload.soe.ucsc.edu/goldenPath) and from ftp://climb.genomics.cn/pub/10.5524/100001_101000/101000/assembly/ (46) for birds. There were 12 females (XX) and 7 males (XY) for mammals, and 10 females (ZW) and 14 males (ZZ) for birds (Datasets S2 and S5). Body mass data are from refs. 103, 106, and 107 for mammals, and from ref. 108 for birds.

For all mammals besides bats, TE annotation was obtained from www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html, RepeatMasker open-4.0.5 (54) ran with the repeat library release 20140131 from repbase (109). For bats and birds, we obtained TE annotations by running RepeatMasker open-4.0.5 (using -e ncbi) with custom libraries (SI Methods).

Determination of Ancient vs. Lineage-Specific TEs. For mammals, we classified TE families as lineage-specific or shared between placental mammals (Dataset S1). We compiled data from Repbase and annotations of the RepeatMasker libraries (54, 109), complemented by our own orthology assessment (combination of BLAT (https://genome.ucsc.edu/FAQ/FAQblat.html), 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 (38, 46). CR1 have been active at least since the common ancestor of birds, always with several subfamilies at the same time (110). This is because one CR1 lineage survived from the many lineages of LINE present in the common ancestor of birds and crocodilians (111). 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 (110)], 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 (parseRM.pl, available at https://github.com/4ureliek/Parsing-RepeatMasker-Outputs) to parse the raw alignment outputs from RepeatMasker (.align files). This process 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 is chosen for that position.

DNA Loss Calculation. DNA loss coefficients were calculated as in Lindblad-Toh et al. (39). 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 S2 for all values and coefficients of other species. For birds, we used $A = 1{,}300$ Mb (SI Methods) and $t = 70$ My [onset of the Neoaves radiation (55)]. We also characterized total loss at the same evolutionary timescales as the ones of our microdeletions and midsize deletion calculations (Fig. S4).

Phylogenetic Correction by Independent Contrasts. To account for phylogenetic dependence (112, 113), we used Felsenstein's independent contrasts method implemented in the PDAP package (114) of Mesquite (115) (SI Methods and Dataset S2). To plot the data in R for Fig. 4, we generated the File of Independent Contrasts and divided the Unstandardized Contrasts of each trait by their SD (Dataset S2).

Analysis of Microdeletions Using a Multispecies Alignment. We developed custom Perl scripts (MAFmicrodel, v2.7), available at https://github.com/4ureliek/MAF_parsing, to recover gaps <30 nt from the MultiZ alignment of human chromosomes 1–22 with 100 other species from the UCSC genome browser (MAF format, hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz100way/). Studied species are listed in Fig. 5, with *M. domestica* and *A. carolinensis* as outgroups for placental mammals and birds, respectively (Dataset S3). Gaps in alignments were placed on a phylogenetic tree based on parsimony and using

intersections of gap coordinates with Bedtools (116). See SI Methods for more details.

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, (v4.6, available at https://github.com/4ureliek/DelGet). See SI Methods for more details.

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 1–22 with 100 other species from the UCSC genome browser (MAF format, hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz100way/). See SI Methods for more details.

1. Petrov DA (2001) Evolution of genome size: New approaches to an old problem. *Trends Genet* 17(1):23–28.
2. Gregory TR (2005) *The Evolution of the Genome* (Elsevier Academic, San Diego, CA).
3. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, MA).
4. Linquist S, et al. (2015) Applying ecological models to communities of genetic elements: The case of neutral theory. *Mol Ecol* 24(13):3232–3242.
5. Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E (2015) Transposons, genome size, and evolutionary insights in animals. *Cytogenet Genome Res* 147(4):217–239.
6. Elliott TA, Gregory TR (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos Trans R Soc Lond B Biol Sci* 370(1678):20140331.
7. Elliott TA, Gregory TR (2015) Do larger genomes contain more diverse transposable elements? *BMC Evol Biol* 15:69.
8. Ellegren H, Galtier N (2016) Determinants of genetic diversity. *Nat Rev Genet* 17(7):422–433.
9. Cavalier-Smith T (1982) Skeletal DNA and the evolution of genome size. *Annu Rev Biophys Bioeng* 11:273–302.
10. Gregory TR, Hebert PD (1999) The modulation of DNA content: Proximate causes and ultimate consequences. *Genome Res* 9(4):317–324.
11. Andrews CB, Mackenzie SA, Gregory TR (2009) Genome size and wing parameters in passerine birds. *Proc Biol Sci* 276(1654):55–61.
12. Wright NA, Gregory TR, Witt CC (2014) Metabolic 'engines' of flight drive genome size reduction in birds. *Proc Biol Sci* 281(1779):20132780.
13. Petrov DA (2002) Mutational equilibrium model of genome size evolution. *Theor Popul Biol* 61(4):531–544.
14. Nam K, Ellegren H (2012) Recombination drives vertebrate genome contraction. *PLoS Genet* 8(5):e1002680.
15. Vicient CM, et al. (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11(9):1769–1784.
16. Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. *Ann Bot (Lond)* 95(1):127–132.
17. Piegu B, et al. (2006) Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 16(10):1262–1269.
18. Vitte C, Panaud O, Quesneville H (2007) LTR retrotransposons in rice (*Oryza sativa, L.*): Recent burst amplifications followed by rapid DNA loss. *BMC Genomics* 8:218.
19. Hawkins JS, Proulx SR, Rapp RA, Wendel JF (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc Natl Acad Sci USA* 106(42):17811–17816.
20. Kelly LJ, et al. (2015) Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytol* 208(2):596–607.
21. Nystedt B, et al. (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497(7451):579–584.
22. Petrov DA, Lozovskaya ER, Hartl DL (1996) High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384(6607):346–349.
23. Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2000) Evidence for DNA loss as a determinant of genome size. *Science* 287(5455):1060–1062.
24. Bensasson D, Petrov DA, Zhang D-X, Hartl DL, Hewitt GM (2001) Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol Biol Evol* 18(2):246–253.
25. Wang X, et al. (2014) The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun* 5:2957.
26. Blass E, Bell M, Boissinot S (2012) Accumulation and rapid decay of non-LTR retrotransposons in the genome of the three-spine stickleback. *Genome Biol Evol* 4(5):687–702.
27. Aparicio S, et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297(5585):1301–1310.
28. Neafsey DE, Palumbi SR (2003) Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes. *Genome Res* 13(5):821–830.
29. Sun C, López Arriaza JR, Mueller RL (2012) Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol Evol* 4(12):1340–1348.
30. Sun C, et al. (2012) LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol Evol* 4(2):168–183.
31. Schubert I, Vu GTH (2016) Genome stability and evolution: Attempting a holistic view. *Trends Plant Sci* 21(9):749–757.
32. Gregory TR (2016) Animal Genome Size Database. Available at www.genomesize.com. Accessed December 12, 2016.
33. Shedlock AM, Edwards SV (2009) Amniotes (Amniota). *The Timetree of Life*, eds Hedges SB, Kumar S (Oxford Univ Press, Oxford), pp 375–379.
34. Meredith RW, et al. (2011) Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334(6055):521–524.
35. Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 10(10):725–732.
36. Waterston RH, et al.; Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520–562.
37. Gibbs RA, et al.; Rat Genome Sequencing Project Consortium (2004) Genome sequence of the brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982):493–521.
38. Hillier LW, et al.; International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018):695–716.
39. Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–819.
40. Gibbs RA, et al.; Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316(5822):222–234.
41. She X, Cheng Z, Zöllner S, Church DM, Eichler EE (2008) Mouse segmental duplication and copy number variation. *Nat Genet* 40(7):909–914.
42. St John J, Quinn TW (2008) Identification of novel CR1 subfamilies in an avian order with recently active elements. *Mol Phylogenet Evol* 49(3):1008–1014.
43. Nicholas TJ, et al. (2009) The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res* 19(3):491–499.
44. Elsik CG, et al.; Bovine Genome Sequencing and Analysis Consortium (2009) The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science* 324(5926):522–528.
45. Bolisetty M, Blomberg J, Benachenhou F, Sperber G, Beemon K (2012) Unexpected diversity and expression of avian endogenous retroviruses. *MBio* 3(5):e00344-12.
46. Zhang G, et al.; Avian Genome Consortium (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346(6215):1311–1320.
47. Chalopin D, Naville M, Plard F, Galiana D, Volff JN (2015) Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol Evol* 7(2):567–580.
48. Ji Y, DeWoody JA (2016) Genomic landscape of long terminal repeat retrotransposons (LTR-RTs) and solo LTRs as shaped by ectopic recombination in chicken and zebra finch. *J Mol Evol* 82(6):251–263.
49. Thomas JW, et al. (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424(6950):788–793.
50. Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV (2007) Origin of avian genome size and structure in non-avian dinosaurs. *Nature* 446(7132):180–184.

51. Hughes AL, Hughes MK (1995) Small genomes for better flyers. *Nature* 377(6548): 391.

52. Zhang Q, Edwards SV (2012) The evolution of intron size in amniotes: A role for powered flight? *Genome Biol Evol* 4(10):1033–1043.

53. Lindblad-Toh K, et al.; Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Genome Institute at Washington University (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7370):476–482.

54. Smit AFA, Hubley R, Green P (2015) RepeatMasker Open-4.0.2013-2015. Available at www.repeatmasker.org. Accessed March 2, 2015.

55. Jarvis ED, et al. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–1331.

56. Blanchette M, Green ED, Miller W, Haussler D (2004) Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res* 14(12):2412–2423.

57. Ma J, et al. (2006) Reconstructing contiguous regions of an ancestral genome. *Genome Res* 16(12):1557–1565.

58. Organ CL, Edwards SV (2011) Major events in avian genome evolution. *Living Dionsaurs: The Evolutionary History of Modern Birds*, eds Dyke GJ, Kaiser GW (John Wiley & Sons, West Sussex, UK), pp 325–337.

59. Team RDC (2008) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

60. Teeling EC, et al. (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307(5709):580–584.

61. Hedges SB, Dudley J, Kumar S (2006) TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971–2972.

62. Ophir R, Graur D (1997) Patterns and rates of indel evolution in processed pseudogenes from humans and murids. *Gene* 205(1-2):191–202.

63. Laurie S, Toll-Riera M, Radó-Trilla N, Albà MM (2012) Sequence shortening in the rodent ancestor. *Genome Res* 22(3):478–485.

64. Bromham L (2011) The genome as a life-history character: Why rate of molecular evolution varies between mammal species. *Philos Trans R Soc Lond B Biol Sci* 366(1577):2503–2513.

65. Sung W, et al. (2016) Evolution of the insertion-deletion mutation rate across the Tree of Life. *G3 (Bethesda)* 6(8):2583–2591.

66. Shedlock AM (2006) Phylogenomic investigation of CR1 LINE diversity in reptiles. *Syst Biol* 55(6):902–911.

67. Shedlock AM, et al. (2007) Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc Natl Acad Sci USA* 104:2767–2772.

68. Janes DE, Organ CL, Fujita MK, Shedlock AM, Edwards SV (2010) Genome evolution in *Reptilia*, the sister group of mammals. *Annu Rev Genomics Hum Genet* 11:239–264.

69. Adams MD, et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287(5461):2185–2195.

70. Gregory TR (2003) Is small indel bias a determinant of genome size? *Trends Genet* 19(9):485–488.

71. Gregory TR (2004) Insertion-deletion biases and the evolution of genome size. *Gene* 324:15–34.

72. Mills RE, et al. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* 16(9):1182–1190.

73. Nóbrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM (2004) Megabase deletions of gene deserts result in viable mice. *Nature* 431(7011):988–993.

74. Han K, et al. (2007) Alu recombination-mediated structural deletions in the chimpanzee genome. *PLoS Genet* 3(10):1939–1949.

75. McLean C, Bejerano G (2008) Dispensability of mammalian DNA. *Genome Res* 18(11):1743–1751.

76. Ponting CP, Nellåker C, Meader S (2011) Rapid turnover of functional sequence in human and other genomes. *Annu Rev Genomics Hum Genet* 12:275–299.

77. Rands CM, Meader S, Ponting CP, Lunter G (2014) 8.2% of the human genome is constrained: Variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet* 10(7):e1004525.

78. Chan YF, et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327(5963):302–305.

79. McLean CY, et al. (2011) Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471(7337):216–219.

80. Hiller M, Schaar BT, Bejerano G (2012) Hundreds of conserved non-coding genomic regions are independently lost in mammals. *Nucleic Acids Res* 40(22):11463–11476.

81. Meredith RW, Zhang G, Gilbert MTP, Jarvis ED, Springer MS (2014) Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* 346(6215):1254390.

82. Lovell PV, et al. (2014) Conserved syntenic clusters of protein coding genes are missing in birds. *Genome Biol* 15(12):565.

83. Chaisson MJ, Wilson RK, Eichler EE (2015) Genetic variation and the de novo assembly of human genomes. *Nat Rev Genet* 16(11):627–640.

84. Lynch M, Bobay LM, Catania F, Gout JF, Rho M (2011) The repatterning of eukaryotic genomes by random genetic drift. *Annu Rev Genomics Hum Genet* 12:347–366.

85. Whitney KD, Boussau B, Baack EJ, Garland T, Jr (2011) Drift and genome complexity revisited. *PLoS Genet* 7(6):e1002092.

86. Figuet E, et al. (2016) Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Mol Biol Evol* 33(6):1517–1527.

87. Szitenberg A, et al. (2016) Genetic drift, not life history or RNAi, determine long term evolution of transposable elements. *Genome Biol Evol* 8(9):2964–2978.

88. Konkel MK, Batzer MA (2010) A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol* 20(4):211–221.

89. Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res* 10(7):908–915.

90. Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12(7):1075–1079.

91. Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14(5):860–869.

92. Tiley GP, Burleigh JG (2015) The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evol Biol* 15:194.

93. van de Lagemaat LN, Gagnier L, Medstrand P, Mager DL (2005) Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res* 15(9):1243–1249.

94. Sen SK, et al. (2006) Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet* 79(1):41–53.

95. Startek M, et al. (2015) Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. *Nucleic Acids Res* 43(4):2188–2198.

96. Lee J, Ha J, Son S-Y, Han K (2012) Human genomic deletions generated by SVA-associated events. *Comp Funct Genomics* 2012:807270.

97. Pagán HJT, et al. (2012) Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. *Genome Biol Evol* 4(4):575–585.

98. Platt RN, 2nd, et al. (2015) Targeted capture of phylogenetically-informative Ves SINE insertions in genus *Myotis*. *Genome Biol Evol* 7(6):1664–1675.

99. Organ CL, Shedlock AM (2009) Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol Lett* 5(1):47–50.

100. Gregory TR, Andrews CB, McGuire JA, Witt CC (2009) The smallest avian genomes are found in hummingbirds. *Proc Biol Sci* 276(1674):3753–3757.

101. Shen Y-Y, Shi P, Sun Y-B, Zhang Y-P (2009) Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. *Genome Res* 19(10):1760–1765.

102. Gregory TR (2002) Genome size and developmental parameters in the homeothermic vertebrates. *Genome* 45(5):833–838.

103. Clarke A, Rothery P, Isaac NJ (2010) Scaling of basal metabolic rate with body mass and temperature in mammals. *J Anim Ecol* 79(3):610–619.

104. Vinogradov AE (1997) Nucleotypic effect in homeotherms: Body-mass independent resting metabolic rate of passerine birds is related to genome size. *Evolution* 51(1):220–226.

105. Smith JDL, Gregory TR (2009) The genome sizes of megabats (Chiroptera: Pteropodidae) are remarkably constrained. *Biol Lett* 5(3):347–351.

106. Garland T (1983) The relation between maximal running speed and body-mass in terrestrial mammals. *J Zool* 199(Feb):157–170.

107. Stuart JA, Page MM (2010) Plasma IGF-1 is negatively correlated with body mass in a comparison of 36 mammalian species. *Mech Ageing Dev* 131(9):591–598.

108. Weber CC, Nabholz B, Romiguier J, Ellegren H (2014) Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol* 15(12):542.

109. Bao W, Kojima KK, Kohany O (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11.

110. Suh A, et al. (2011) Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun* 2:443.

111. Suh A, et al. (2014) Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol Evol* 7(1):205–217.

112. Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125(1):1–15.

113. Garland T, Jr, Bennett AF, Rezende EL (2005) Phylogenetic approaches in comparative physiology. *J Exp Biol* 208(Pt 16):3015–3035.

114. Midford PE, Garland T, Maddison WP (2005) PDAP Package of Mesquite. Version 1.07. Available at https://github.com/MesquiteProject/Mesquite_PDAP. Accessed December 5, 2016.

115. Maddison WP, Maddison DR (2006) Mesquite: A modular system for evolutionary analysis. Version 1.1. Available at mesquiteproject.org. Accessed December 5, 2016.

116. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

117. Pace JK, 2nd, Gilbert C, Clark MS, Feschotte C (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci USA* 105(44):17023–17028.

118. Ray DA, et al. (2008) Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res* 18(5):717–728.

119. Mitra R, et al. (2013) Functional characterization of *piggyBat* from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc Natl Acad Sci USA* 110(1):234–239.

120. Thomas J, Phillips CD, Baker RJ, Pritham EJ (2014) Rolling-circle transposons catalyze genomic innovation in a mammalian lineage. *Genome Biol Evol* 6(10):2595–2610.

121. Zhuo X, Feschotte C (2015) Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PLoS Pathog* 11(11):e1005279.

122. Eöry L, et al. (2015) Avianbase: A community resource for bird genomics. *Genome Biol* 16:21.

123. Warren WC, et al. (2010) The genome of a songbird. *Nature* 464(7289):757–762.

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

125. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.

EVOLUTION