Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *The Year in Evolutionary Biology*
REVIEW ARTICLE

Evolution of bird genomes—a transposon's-eye view

Aurélie Kapusta[1] and Alexander Suh[2]

[1]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah. [2]Department of Evolutionary Biology (EBC), Uppsala University, Uppsala, Sweden

Address for correspondence: Alexander Suh, Department of Evolutionary Biology (EBC), Uppsala University, Uppsala SE-75236, Sweden. alexander.suh@ebc.uu.se.

Birds, the most species-rich monophyletic group of land vertebrates, have been subject to some of the most intense sequencing efforts to date, making them an ideal case study for recent developments in genomics research. Here, we review how our understanding of bird genomes has changed with the recent sequencing of more than 75 species from all major avian taxa. We illuminate avian genome evolution from a previously neglected perspective: their repetitive genomic parasites, transposable elements (TEs) and endogenous viral elements (EVEs). We show that (1) birds are unique among vertebrates in terms of their genome organization; (2) information about the diversity of avian TEs and EVEs is changing rapidly; (3) flying birds have smaller genomes yet more TEs than flightless birds; (4) current second-generation genome assemblies fail to capture the variation in avian chromosome number and genome size determined with cytogenetics; (5) the genomic microcosm of bird–TE "arms races" has yet to be explored; and (6) upcoming third-generation genome assemblies suggest that birds exhibit stability in gene-rich regions and instability in TE-rich regions. We emphasize that integration of cytogenetics and single-molecule technologies with repeat-resolved genome assemblies is essential for understanding the evolution of (bird) genomes.

Keywords: bird; transposable element; endogenous virus; genome evolution; chromosome; long-read sequencing

The uniqueness of bird genomes

Birds are the only surviving dinosaurs[1] and played an important role in Charles Darwin's contribution to evolutionary biology.[2] In the meantime, birds have become one of the most intensively studied groups of organisms in many fields, such as research on biogeography,[3,4] development,[5,6] domestication,[7] neurobiology,[8,9] phylogenetics,[10–13] and speciation.[14–18] Even more recently, birds have become one of the most densely sampled higher-level animal taxa in genomics research.[19] From 2004 until 2010, the first sequenced bird genomes were the agriculturally important chicken[20] and turkey,[21] as well as the zebra finch,[22] a model system for vocal learning. Since 2012 and resulting from the massive reductions in sequencing costs and effort spurred by high-throughput short-read sequencing, many laboratories have pursued bird genome projects individually (e.g., Refs. 14, 16, 17, and 23–25) and as part of the Avian Phylogenomics Consortium (e.g., Refs. 5, 26, and 27). As of June 2016, a total of 76 genome assemblies from 75 bird species have been published or made publicly available in GenBank (Table S1, online only). The Bird 10,000 Genomes project is currently aiming to sequence the remainder of the ~10,500 extant bird species.[28] This places birds as an ideal case study for genome evolution, especially in the light of recent developments in genomics research.

Despite these extensive recent and ongoing efforts, an integrated perspective on avian genome evolution is lacking. This is particularly true for noncoding regions of the genome, many of which harbor repetitive genomic parasites, such as transposable elements (TEs) and endogenous viral elements (EVEs). TEs are selfish genetic elements that either copy and paste via an RNA intermediate (i.e., class I elements or retrotransposons) or directly cut and paste in their DNA form (i.e., most class II elements or DNA transposons).[29,30] EVEs are relics of virus infections that occurred in the

doi: 10.1111/nyas.13295

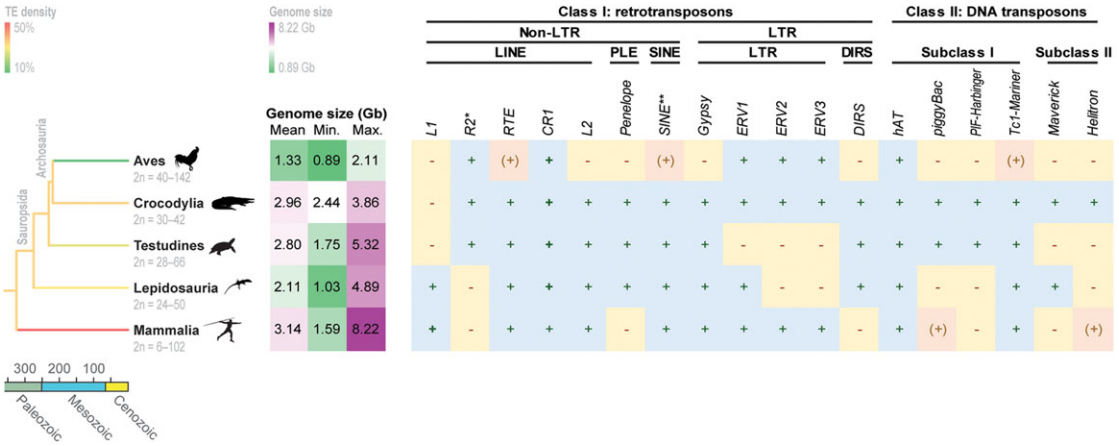

Figure 1. The diversity of transposable elements in amniotes. Evolution of TE density across the phylogeny of land vertebrates[187] in comparison with karyotype ranges[58] (diploid chromosome counts in gray), a heat map of genome size ranges (data from Ref. 47 converted into Gb[188]), and a distribution table of TE superfamilies (*sensu* Kapitonov and Jurka[38] and Wicker *et al.*[37]). The TE distribution table follows Kordiš,[62] updated with our analysis of publicly available RepeatMasker annotations (http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html) of one representative genome per major amniote lineage, namely chicken,[20] American alligator,[94] painted turtle,[189] anole lizard,[190] and human.[90] "+" indicates the presence of a TE superfamily, and "–" indicates spurious short hits or the complete absence of a TE superfamily. To give a more complete view of the TE diversity in each lineage, "(+)" indicates the presence of a TE superfamily in few host subtaxa due to recent horizontal transfer or *de novo* emergence. The asterisk denotes the widespread presence of R2 in low copy numbers;[88] the double asterisk indicates that ancient CORE-SINEs are present in all amniote genomes.[91–93] Bold "+" indicates dominance of this TE superfamily in the host lineage.

host germline and thereby became vertically transmitted over long evolutionary timescales.[31] The paleovirological record of known genomic "fossils" of viruses dates back to > 207 Mya[32] and spans all major groups of eukaryotic viruses (reviewed in Refs. 31 and 33–35). However, the most commonly found EVEs are retroviruses that rely on obligate integration into the host genome;[35,36] these EVEs can also be classified as long terminal repeat (LTR) retrotransposons.[37,38] There is thus no clear boundary between the world of transposons ("transpososphere") and the world of viruses (virosphere; see also Ref. 39). Furthermore, these genomic parasites are present in virtually all cellular organisms[40] and influence the evolution of their hosts on multiple levels, such as genome size,[41] transcriptional regulation,[42,43] 3D genome folding,[44,45] and many types of structural variation (reviewed by, e.g., Refs. 29 and 46).

Here, we provide an integrated perspective on avian TEs and EVEs and their impact on avian genome evolution, and illuminate future avenues of avian genomics research once repeat-resolved third-generation genome assemblies are available with in-depth repeat annotations.

Why are bird genomes unique?

Birds have the smallest genomes among amniotes.[47] The variation in bird genome size is relatively low (Fig. 1),[48] ranging from 0.89 Gb in the black-chinned hummingbird[49] to 2.11 Gb in the ostrich.[47] Although a significant reduction in genome size (and TE density) began after the bird/crocodilian split and probably predated the evolution of flight,[50] avian genome size has generally continued to decrease since the common ancestor of birds.[48,51] There is now ample evidence for constraints on cell and genome sizes due to the metabolic requirements of powered flight,[49,51–54] but the exact links are still debated and might be associated with DNA repair mechanisms.[55] Consistent with the proportional model of genome evolution of Oliver *et al.*,[56] which assumes less evolutionary change in smaller genomes (and more evolutionary change in larger genomes), analyses of the first genome assemblies have led to the conclusion that bird genomes have unusually stable chromosomes.[57] However, cytogenetic data suggest that birds have the highest variance in chromosome numbers among amniotes (Fig. 1).[58] The high chromosome numbers of birds are due to the presence of many

small chromosomes (microchromosomes, most of them considerably smaller than 20 Mb[20,59]), a genomic feature that emerged in the avian ancestor due to fission of ancestral macrochromosomes and medium-sized chromosomes into even smaller microchromosomes.[59,60] In contrast to avian macrochromosomes, microchromosomes exhibit high rates of meiotic recombination, high GC content, small introns, high densities of genes and CpG islands, and low densities of TEs and other repeats.[48,57,59] Just as bird genomes are more streamlined than other amniote genomes, their genomic properties suggest that avian microchromosomes are even more streamlined than the rest of the avian genome. Efficient selection due to high recombination rates (reviewed in Ref. 61) may explain the evolution and maintenance of small, repeat-poor genomes and chromosomes, respectively.[59]

How do the properties of the avian genome affect the prevalence of repetitive genomic parasites such as TEs and EVEs? In contrast to nonavian reptiles and mammals (Fig. 1 and reviewed in Refs. 46 and 62–65), birds have low densities of TEs and EVEs. Genome assemblies are typically 1.0–1.3 Gb in size and contain between 130,000 and 350,000 TE copies and a TE density of 4.1–9.8%[20,22,26] (but note the downy woodpecker with ∼700,000 TEs making up 22.2% of the genome[26]). Furthermore, birds also have a low diversity of TE superfamilies (see also Ref. 26). Only a few of the TE superfamilies widespread in amniotes are present in all birds (Fig. 1). From these, the dominant TE superfamily in birds is chicken repeat 1 (CR1),[20,22,26] which dominates amniote genomes, with the exception of the L1-dominated therian mammals and the L2-dominated monotreme mammals (Fig. 1).[66] This CR1 dominance has been suggested to reflect the genome organization of the amniote ancestor[64,66,67] and explains the scarcity of retro(pseudo)genes in the virtually L1-lacking bird genomes.[20,68] More precisely, most bird genomes contain only one ancient full-length L1 copy or remnants thereof.[69]

To sum up, the body of avian genomics research suggests that bird genomes are generally small, stable, and devoid of TEs. But is that it, really? In the following, we argue that this picture is an oversimplification that arose from the restriction of in-depth genome analyses to model organisms (chicken and zebra finch) and limitations of short-read sequencing technologies. As outlined below, in-depth analyses of dozens of bird genomes revealed a previously hidden diversity of TEs and EVEs. Furthermore, upcoming long-read genome assemblies (hereafter referred to as "third-generation genome assemblies") contain many repeat-rich genomic regions that are largely invisible in current short-read genome assemblies (hereafter referred to as "second-generation genome assemblies").

The diversity of avian transposons and endogenous viral elements

LINE retrotransposons

Long interspersed elements (LINEs) are autonomous retrotransposons without LTRs (i.e., non-LTR retrotransposons). They usually contain two open reading frames (ORFs) coding for the ORF1 protein and a reverse transcriptase (RT) protein with an endonuclease domain (Fig. 2).[37,38] Intact and transcribed LINE master genes retrotranspose via a common mechanism,[70] target-primed reverse transcription (TPRT).[71] While the ORF1 protein and the LINE mRNA form a ribonucleoprotein particle, the RT protein nicks the genomic target site and reverse-transcribes the LINE mRNA into cDNA (reviewed in Refs. 72 and 73). The nicking of the target site usually occurs via staggered cuts, leading to the emergence of a target site duplication (TSD) of variable length in the resultant LINE daughter copy (Fig. 2).[37] Daughter copies are often 5′-truncated due to premature termination of TPRT, which prevents them from being transcribed and retrotransposed.[73]

CR1 elements are by far the most abundant TEs in birds (39–88% of all TE copies)[20,22,26] and were the first TEs to be discovered in any nonmammalian amniote.[74,75] It has since become clear that the CR1 superfamily is widespread in animals,[76] includes the mammalian L3 elements,[66,68] and dominates most amniote genomes.[64,66,67] In birds, at least 14 CR1 families have been described,[20,22] many of which are ancient.[77] They all diversified from a common CR1 ancestor in early bird evolution, whereas the six other ancient CR1 lineages present in the bird/crocodilian ancestor went extinct.[66,68] Many of these diverse CR1 families temporally overlapped in activity, were active over long periods (cf. Fig. 3) and are thus useful phylogenetic markers for reconstructing various avian relationships.[78–82] Nearly all genomic copies of CR1 are heavily truncated at their 5′ ends[20,77,83] and thus lack CR1 promoters or coding

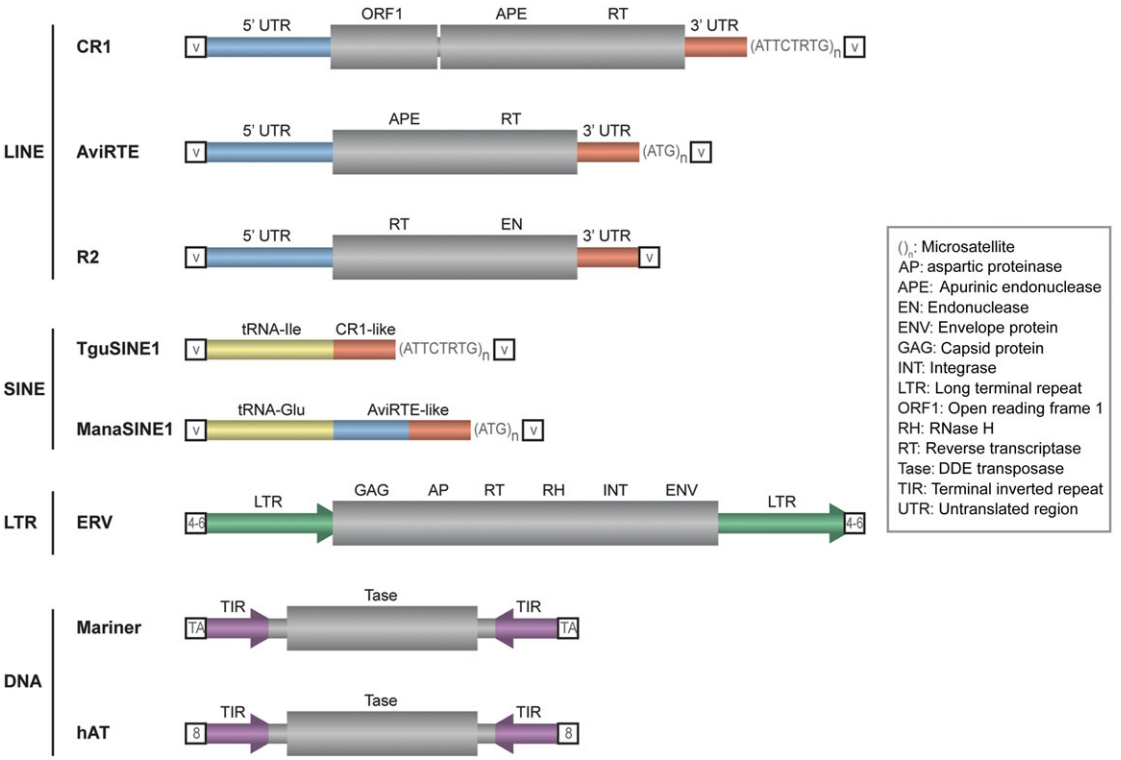

Figure 2. **The major superfamilies of avian transposable elements. Schematic organizations of TE superfamilies after Refs. 37, 84, and 95 and following the classification shown in Figure 1. CR1 is the dominant LINE superfamily of birds and is structurally very similar to L1 (dominant in therian mammals and virtually absent in birds (Fig. 1)), except for the presence of a hairpin and octamer microsatellite tail instead of an A-rich tail.[68] Structural elements are color coded, namely 5′ UTRs (blue), 3′ UTRs (red), LTRs (green), ORFs (gray), TIRs (purple), and tRNAs (yellow). White squares are target site duplications, which are either of variable (v) or of specific length (as indicated by numbers and letters). Gray letters in parentheses indicate microsatellite motifs.**

capacity. However, they all contain the CR1 3′ UTR with the typical hairpin and octamer microsatellite motifs (Fig. 2),[68] and it remains unknown whether these structures influence the stability and transcriptional regulation of adjacent genomic regions.

Recent in-depth analyses of bird genomes have identified additional, peculiar LINE superfamilies (Fig. 2), albeit in much lower copy numbers than CR1. There is now evidence for repeated horizontal transfer (HT) of AviRTE because of its presence and recent activity in unrelated bird lineages, namely suboscine passerines, psittacid parrots, hornbills, trogons, hummingbirds, mesites, and tinamous.[84] Notably, the presence of AviRTE in some filarial nematodes implies that these insect-borne endoparasites were potential vectors of HT. AviRTE belongs to the RTE superfamily, just like BovB, which was previously shown to have undergone widespread, tick-mediated HT between some mammals and reptiles.[85,86] Furthermore, the R2 superfamily is widespread among animals and integrates specifically into 28S rRNA genes.[87] In birds, R2 has been described only from the zebra finch;[62] but very recent in-depth annotations suggest that low copy numbers of R2 elements are present in most bird genome assemblies spanning the entire avian tree of life.[88]

SINE retrotransposons

Short interspersed elements (SINEs) are nonautonomous non-LTR retrotransposons that parasitize the enzymatic machinery of LINEs. They lack protein-coding capacity and thus undergo *trans*-mobilized TPRT. This is achieved via sequence similarity between SINE tails and LINE 3′ UTRs that leads to recognition of the SINE RNA by the LINE RT protein.[89] SINE heads are usually derived from

tRNA, 5S rRNA, or other small RNA genes, and these internal promoters permit transcription by RNA polymerase III.[37,38] SINEs are thus part of the small RNA transcriptome.

Generally, birds exhibit low copy numbers of SINEs (6000–17,000[26] versus >1,500,000 in humans[90]) and the majority of them are ancient, heavily degraded elements from the L2-mobilized MIR.[20,26] Additional ancient SINEs with moderate copy numbers are AmnSINE and LFSINE,[26] which belong to the group of CORE-SINEs and are present in all amniote genomes.[91–94] Similar to MIR, these SINEs exhibit L2-like tails, suggesting comobilization by L2. Insights from mammalian genomes suggest that some of these SINEs contain insulator motifs and are involved in 3D genome folding,[44,45] which may explain the visibility of such ancient SINEs in all amniote genomes. As all of these SINEs have been inactive for very long times, the initial analysis of the chicken genome suggested that birds lack any activity of SINEs.[20]

However, there is now a growing body of evidence suggesting recurrent *de novo* emergence and recent activity of SINEs in some avian lineages (see Fig. 4 for an overview). CR1-mobilized SINEs contain the aforementioned structures typical for CR1 3′ UTRs,[68] tRNA-derived heads (Fig. 2), and were first discovered in the zebra finch genome.[22] This particular SINE, TguSINE1, was recently shown to have emerged in the common ancestor of oscine and suboscine passerines and, via template switching, given rise to a CR1-SINE with a different tRNA-derived head in suboscines.[95] Further instances of *de novo* emergence of CR1-SINEs were recently detected in the genomes of the pelican *Pelecanus crispus* and the trogon *Apaloderma vittatum* (A. Suh, unpublished data). Additionally, there is a striking diversity of RTE-mobilized SINEs with different heads (i.e., derived from tRNA, 5S rRNA, 28S rRNA, or unknown GC-rich sequence) and AviRTE-derived bipartite tails (Fig. 2; see also fig. 1 of Ref. 84). Although AviRTE underwent extensive HT between birds, their RTE-SINEs appear to be lineage-specific emergences, namely three in suboscine passerines, one in psittacid parrots, and one in hornbills.[84] This situation is again reminiscent of the aforementioned RTE family BovB, this time because of mobilization of a similar diversity of SINEs in mammals and reptiles.[96]

LTR retrotransposons

LTR retrotransposons mobilize via replicative retrotransposition, a mechanism that is strikingly different from the TPRT mechanism of LINEs (e.g., reviewed by Refs. 29 and 73). The diversity of LTR retrotransposons includes endogenous and exogenous retroviruses (ERVs and XRVs), and autonomous elements exhibit protein-coding capacity for a variety of viral proteins (Fig. 2).[37] Reverse transcription of LTR mRNA is primed by a host tRNA[29] and occurs within virus-like particles in the cytoplasm.[73] The resultant full-length double-stranded cDNA is then transported into the nucleus and inserted into the genome via the integrase protein,[73] leading to TSDs of 4–6 bp in length (Fig. 2).[37] Each daughter copy is thus a full-length LTR retrotransposon,[73] unlike LINEs and SINEs which have frequent 5′-truncation during TPRT. However, the high sequence similarity between the left and right LTRs of a single transposon frequently leads to intra-element ectopic recombination, a mechanism of genome size contraction following LTR amplifications.[97] This entails the removal of the internal protein-coding regions and one of the two LTRs, leaving behind a solo-LTR flanked by the original TSD.[97] Although unable to mobilize, solo-LTRs contain promoters and thus affect transcriptional regulation of nearby genes.[98]

In birds, most LTR retrotransposon copies constitute solo-LTRs.[83] Full-length elements are rare, and analyses of their protein-coding sequences suggest that most avian LTR retrotransposons are (endogenous) retroviruses, predominantly alpharetroviruses and gammaretroviruses[99] (*contra* Ref. 65 reporting dominance of betaretroviruses and gammaretroviruses). Interestingly, in-depth ERV analysis of the zebra finch genome revealed a recombinant retrovirus that acquired its envelope protein gene from a mammalian gammaretrovirus and underwent recent cross-species transmissions between oscine passerines, woodpecker, and duck.[100] On the other hand, analysis of the higher copy number solo-LTRs have yielded insights into LTR accumulation across the avian tree of life. Presence/absence analyses of orthologous solo-LTRs suggest that LTR activity occurred during the early evolution of neognaths (Neoaves and Galloanserae).[12,79,101] Generally, available in-depth TE annotations imply higher LTR accumulation in the zebra finch lineage than in the chicken lineage.[20,22,83] This is partly due to

an expansion of the LTR activity and diversity during the neoavian radiation.[12,79] However, most LTR accumulation appears to have occurred relatively recently in the highly species-rich oscine passerines (e.g., zebra finch, collared flycatcher, and American and hooded crows) because of diversification of the LTR superfamilies ERV1, ERV2, and ERV3 (cf. Fig. 3).[22,65,102,103] Recent in-depth annotations of collared flycatcher[102] and hooded crow[103] revealed a high diversity of lineage-specific LTR retrotransposons, which coincides with a reduction in recent CR1 activity.[46,102] Very recent data imply that these patterns of LTR dominance and potential CR1 extinction probably emerged independently after the massive diversification of oscine passerines (A. Suh, unpublished data).

DNA transposons

In contrast to the aforementioned copy-and-paste retrotransposons, DNA transposons move via a cut-and-paste mechanism that requires a transposase protein (see Refs. 37, 38, and 104 for notable exceptions). Typical structural hallmarks of these elements are terminal inverted repeats (TIRs) (Fig. 2), which are bound by the transposase protein, mediating precise DNA cleavage at both transposon termini and reinsertion of the transposon DNA into a new genomic locus (reviewed by Refs. 30, 73, and 104). Transposition usually occurs into specific target motifs and results in well-defined TSDs (Fig. 2).[37] The mechanism of DNA transposon movement is thus dependent on TIR recognition and, consequently, similar to the previously discussed SINE–LINE parasitism, nonautonomous DNA transposons exist. These nonautonomous DNA transposons lack protein-coding capacity owing to internal deletions or rearrangements and therefore rely on *trans*-mobilization by the transposase protein of autonomous DNA transposons.[104]

To date, the only available in-depth analyses of avian DNA transposons are restricted to the chicken genome.[83] They revealed relatively recent activity of Galluhop transposons from the mariner superfamily and Charlie12 transposons from the hAT superfamily (Fig. 2). The majority of these elements are either nonautonomous elements or a peculiar Charlie–Galluhop fusion element, while only a few copies are transposase-encoding autonomous DNA transposons.[83] Furthermore, annotation of the zebra finch genome suggests scarcity of DNA transposons except for a relatively old hAT element, Charlie7, which is present in low copy numbers in both chicken and zebra finch.[22] To investigate the timing of the DNA transposon expansion on the chicken lineage, we dated the activities of TEs in 42 bird genomes (Fig. 3). The hAT and mariner expansion probably occurred in the chicken/turkey ancestor, and we detected no significant amounts of known DNA transposons in any other bird lineage (Fig. 3). Nevertheless, in-depth DNA transposon annotation of most bird genome assemblies is pending, and there is recent evidence for the presence of Galluhop transposons in hornbills due to HT between hornbills and the chicken/turkey lineage (Gabriel L. Wallau, unpublished data).

Endogenous viral elements

Retroviruses are reverse-transcribing ssRNA viruses with LTRs, and retroviral EVEs are present in all analyzed bird genomes,[65] as discussed above. Among non-retroviral EVEs in amniotes, strikingly low copy numbers have been noted in a study mining 48 bird genomes for EVEs from ssDNA viruses (Circoviridae, Parvoviridae), ssRNA viruses (Bornaviridae), and reverse-transcribing dsDNA viruses (Hepadnaviridae).[65] However, it remains unclear whether this scarcity of EVEs results from a rarity of germline infiltrations. Alternatively, we argue that the aforementioned high efficacy of selection in the streamlined bird genomes may lead to low fixation probabilities and rapid turnover of TE and EVE insertions in many genomic backgrounds.

In any case, the EVEs of most virus families have patchy phylogenetic distributions in birds (see Fig. 1 of Ref. 65). Indeed, circovirus EVEs are present in the genomes of medium ground finch, kea, egret, and tinamou; parvovirus EVEs inserted into the genomes of most passerines, egret, pelican, hoatzin, bustard, mesite, and sandgrouse; and bornavirus EVEs are restricted to the genomes of woodpecker, hummingbird, and swift. However, it remains unknown whether the presence of some of these EVEs in close relatives is due to orthologous insertion in their respective common ancestor or to independent germline infiltrations. On the other hand, hepadnavirus EVEs have been reported from all analyzed bird genomes except for Galliformes (chicken and turkey).[31,65,105–108] They exhibit higher copy numbers than other

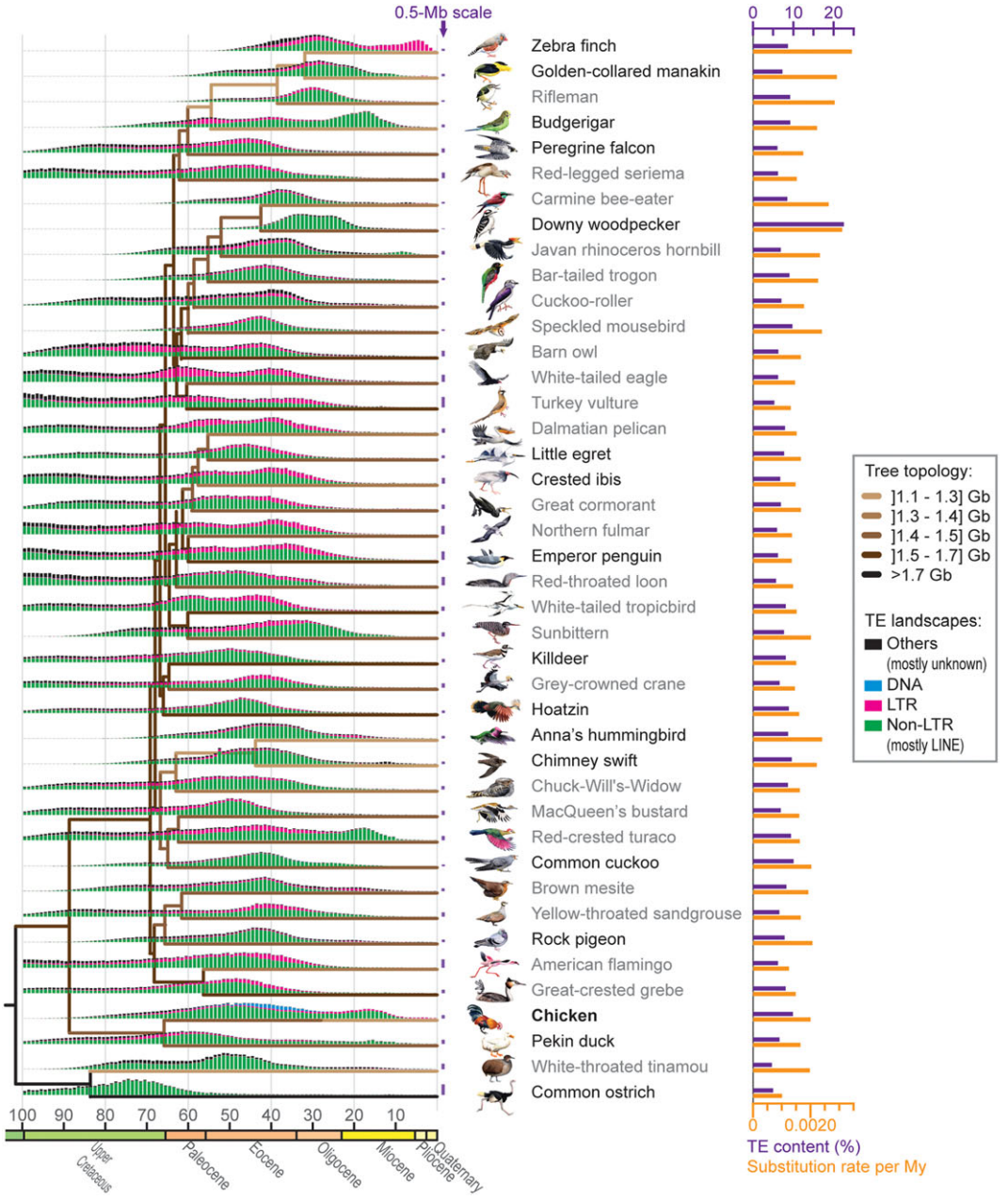

Figure 3. The temporal landscape of transposition across the avian tree of life. The major bird lineages of the Jarvis *et al.*[10] phylogeny are shown with dated TE landscapes as stacked histograms (data from Ref. 113), TE content (purple bars, values from Ref. 113), and substitution rates (orange bars, values from Ref. 26). We note that because the dated TE landscapes rely on extant substitution rates, the ages of older TEs might be incorrectly estimated in bird lineages with possible recent substitution rate shifts (e.g., zebra finch and downy woodpecker). The phylogenetic tree is color coded based on estimated genome sizes in Gb (genome size data from Ref. 47 and extrapolations from assembly sizes/coverage from Ref. 113, combined reconstruction using parsimony as in Ref. 48). TE landscapes are in bins of 1 My and were obtained from the RepeatMasker[162] align outputs (data from Ref. 113 except for the new chicken galGal5 assembly, which was masked with the option "-species Aves" for this review) and each species' substitution rates,[26] processed together with a custom Perl script (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs/blob/master/parseRM.pl). The *y*-axis scale of each histogram is shown as a 0.5-Mb dark purple bar. Common names of the species are grayscale coded according to assembly quality (see Table S1, online only, for more details). The color change on the chicken lineage indicates the timing of the chicken/turkey divergence.

non-retroviral EVEs,[65] and presence/absence analyses of some hepadnavirus EVE orthologs have shown that germline infiltrations occurred throughout avian evolution, from the neoavian ancestor to recent divergences within estrildid finches.[105] Altogether, there is now direct paleovirological evidence for coexistence of birds and hepadnaviruses for >70 My[105] and indirect evidence for deep vertebrate roots of hepadnaviruses.[32] This is in stark contrast to previous studies on extant viruses alone, which estimated the divergence of avian and mammalian hepadnaviruses as recently as 30,000[109] or 125,000[110] years ago. EVEs from bird genomes have thus provided unique windows into the deep evolutionary past of medically relevant viruses.

The temporal landscape of transposition in birds

As mentioned above, chicken and turkey are the only birds with large amounts of DNA transposons in their genomes (visible in blue in Fig. 3). Moreover, oscine passerines (e.g., zebra finch) show a unique recent expansion of LTR retrotransposons (Fig. 3). This illustrates that there is not only large-scale variation in TE diversity in birds, but also a diversity in temporal landscapes of TE accumulation. The chicken and zebra finch genomes exhibit the most recent and abundant TE activity and are also the best assemblies to date (Table S1, online only), suggesting that the amount of recent TE activity in the remaining, mostly Illumina-based, second-generation assemblies is underestimated (Fig. 3). Nevertheless, because old TEs are much more likely to be present even in the assemblies with short reads,[111,112] between-species differences in their overall amounts and landscape shapes are likely due to biological reasons. The most striking example for this is the downy woodpecker, with a massive expansion of CR1 retrotransposons that probably began after the woodpecker/bee-eater split (Fig. 3).

The "accordion" model of genome size evolution

Variation in TE content and temporal landscape between genomes (Fig. 3) does not simply reflect the variation in transposition rates of different TEs at different time points. Rather, TE accumulation is also determined by the rate of fixation of new TE insertions through genetic drift, as well as the degradation or loss of fixed TE copies through fixed nucleotide substitutions and deletions.[113] The latter depends on the age of the TE insertion, as well as nucleotide substitution and deletion rates. In neutrally evolving DNA, these rates depend on the strength of genetic drift because of effective population size associated with life history traits (e.g., body size, generation time).[61,114] Consequently, there is a positive correlation between the rates of substitution and small deletion in birds.[94,113] However, the degree of genomic instability resulting from TE accumulation has been shown to additionally affect deletion rates.[97] Indeed, rapid TE expansion leads to frequent ectopic recombination between similar TE copies, thereby increasing the frequency of medium and large deletion events.[97,115,116]

A recent comparative genomic study investigated the amounts of DNA gained (through TE insertions) and lost (through deletions) during avian evolution.[113] Interestingly, while birds with low TE densities underwent few deletion events, birds that accumulated more TEs also deleted more DNA over the same time frame.[113] In other words, the overall amounts of DNA gained and lost correlate positively in birds. This may explain why the downy woodpecker has an assembly size comparable to the remaining birds, despite a highly increased TE density of 22% (Fig. 3). Comparable observations have been made for mammals, and these led to a unified accordion model of genome size evolution in amniotes[113] (note that a similar metaphor was previously used for "the accordion model of *Mhc* evolution"[117]). Accordingly, genomic instability following TE expansion increases the frequency of deletions and thereby counteracts DNA gains. Some bird genomes are thus more dynamic than their constrained small genome sizes may suggest.

Smaller flyers with faster genomes

Surprisingly, despite having smaller genomes than flightless birds, flying birds generally have higher rates of nucleotide substitution and deletion and higher TE densities (Fig. 3; cf. ostrich and penguin versus remaining birds).[113] This is in line with the aforementioned accordion model——birds with higher TE density have overall fewer old TEs owing to higher deletion and substitution rates, as well as more young TEs owing to higher insertion rates.[113] Conversely, the genomes of flightless birds are larger

as the result of an evolutionary slowdown (i.e., near absence of recent TE insertions and reduced shrinkage of the genome via deletions).[113] This adds further support to a largely nonadaptive mode of genome size evolution[118,119] where TE expansions lead to genome instability, which in turn increases the frequency of deletion events.[113] The resulting indirect maintenance of genome size may help meet the constraints on genome size in birds with powered flight.[49,51–54] Conversely, the potentially relaxed constraints on genome size in secondarily flightless birds did not translate into TE expansions in penguins and ostrich (Fig. 3). This is surprising and requires the study of additional flightless birds.

It is puzzling that the genomes of flying birds are generally more dynamic than those of flightless birds. Because the sampled flightless birds are larger bodied with longer generation times than most of the sampled flying birds,[114] part of this phenomenon likely results from slower genetic drift due to longer generation times, similar to what has been proposed for the slow rates of molecular evolution in crocodilians and turtles.[94] Additionally, we speculate that the more dynamic genomes of flying birds may at least partially result from frequent DNA damage or breakage because of the metabolic stress of powered flight. In this context, we note that the rate of nucleotide substitution has previously been proposed to correlate with species richness in birds.[26,120] Similarly, the aforementioned high density and diversity of LTR retrotransposons in oscine passerines might hint at potential links between TE diversification, TE amplification, genome instability, and species diversification in birds. Which came first, the chicken or the transposon?

Current status and limitations of avian genomics

Our compendium of evolutionary events in avian TEs (Fig. 4) is likely just the tip of the iceberg, as novel aspects of TE evolution are unearthed with every in-depth analysis of available genomes. Where do we go from here? Repetitive elements (including TEs and EVEs) pose one of the major obstacles in assembling genomes, irrespective of the sequencing technology (reviewed in Refs. 111 and 112). Below, we review the current status of avian genome assemblies from the perspective of repetitive elements, genome size, and chromosome number.

Genome assembly quality

A common statistic to assess the quality of genome assemblies is the N50 value (i.e., the minimum length of long sequences that make up half of the assembly, of contigs (contiguous sequences) or scaffolds (linked contigs separated by gaps)).[121] Here, we define genome assemblies as low quality (LQ), medium quality (MQ), high quality (HQ), and gold quality (GQ) according to their overall degree of linkage and contiguity (see also legend of Fig. 4). Under this strict definition, the available 76 avian genome assemblies comprise one GQ, four HQ, 32 MQ, and 36 LQ assemblies (Fig. 4), as well as three contig-level assemblies (Table S1, online only). Another way to assess the quality of a genome assembly is to look at the representation of highly similar repetitive regions (e.g., very young TE insertions), which are difficult to assemble.[130,131] As mentioned above, short-read genome assemblies consistently underestimate the amount of very recent TE-derived DNA (Fig. 3). All of these are MQ or LQ assemblies. On the other hand, the HQ and GQ assemblies contain larger amounts of young TEs (Fig. 3) (e.g., chicken and zebra finch) and exhibit higher contiguity due to classical Sanger sequencing or PacBio long-read sequencing (Table S1, online only).

Genome size variation in birds

Birds have the lowest range of genome sizes among amniotes.[48] However, flow cytometric evidence suggests nearly twofold genome size variation across and even within some bird orders (Fig. 4) (e.g., Anseriformes, Piciformes, and Psittaciformes).[47] Yet, existing genome assemblies all range from 1.0 to 1.3 Gb (Table S1, online only), even in cases with much larger predicted genome sizes. The most striking example is the ostrich, with an Illumina-based second-generation assembly size of 1.23 Gb[27] and a predicted genome size of 2.11 Gb.[47] It is possible that part of this discrepancy arises from the fact that the predicted genome size of the ostrich is based on a single flow cytometric measurement.[47] Nevertheless, we are certain that second-generation assemblies consistently underestimate genome size variation because they yield homogenized assembly sizes, possibly due to issues of short-read sequencing with the recovery of structurally complex and repeat-rich regions (e.g., reviewed by Refs. 111 and 112). If this is indeed the case, then most of the genome size variation in birds is currently

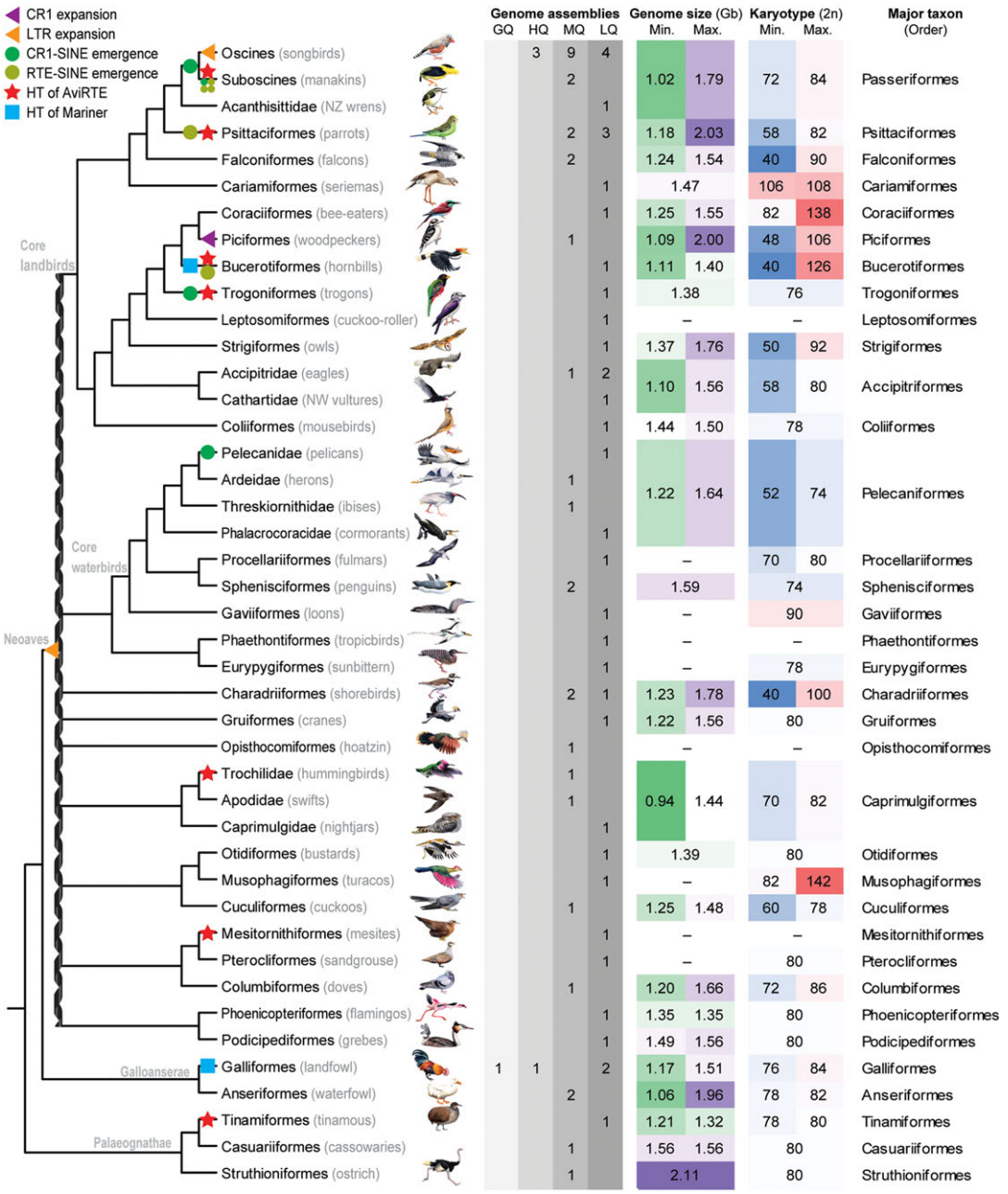

Figure 4. The evolution of transposable elements, genome size, and chromosome number across the avian tree of life. The consensus avian tree of life (after Ref. 13) contains a polytomy at the root of Neoaves due to phylogenetic uncertainty in genome-scale phylogenies.[10–12] For the illustrated major bird lineages (branches with bird paintings are identical to Fig. 3), we contrasted TE evolution with the quality of available genome assemblies (cf. Table S1, online only), a heatmap of genome size ranges (data from Ref. 47), and a heat map of chromosome number ranges (diploid karyotypes, data from Refs. 47, 48, 58, and 123–127). Key events of TE evolution in the avian tree of life are CR1 expansions in woodpeckers,[26] LTR expansions during the neoavian radiation[12,79] and in oscines,[102,103] CR1-SINE emergences in various birds (Ref. 95 and A. Suh, unpublished data), horizontal transfers of AviRTEs and RTE-SINE emergences in various birds,[84] and horizontal transfers of mariner DNA transposons (Ref. 83 and Gabriel L. Wallau, unpublished data). Genome assembly qualities are low quality (LQ, scaffold N50 $<$ 1 Mb), medium quality (MQ, scaffold N50 $>$ 1 Mb), high quality (HQ; chromosome-level assembly with contig N50 $<$ 1 Mb), and gold quality (GQ, chromosome-level with contig N50 $>$ 1 Mb).

inaccessible and may be largely due to unassembled TEs, EVEs, and other repetitive elements, such as tandem repeats in centromeres and telomeres. We anticipate that at least some of the genome size variation will become visible in third-generation genome assemblies of birds owing to improved representation of repetitive elements via long-read sequencing.

Chromosome number variation

Although it has been shown that two-thirds of all bird species have a diploid karyotype of ~80 chromosomes,[122] birds have the largest range of diploid chromosome numbers among amniotes ($2n$ = 40–142; Fig. 1).[47,48,58,123–127] Part of this discrepancy may result from the fact that the only available *de novo* chromosome-level assemblies are a nonrepresentative sample of five birds—chicken[20] and turkey[21] from Galliformes and collared flycatcher,[17] great tit,[25] and zebra finch[22] from oscine Passeriformes. These two avian orders contain nearly two-thirds of all extant bird species and have rather narrow ranges of chromosome numbers (Fig. 4). The high levels of interchromosomal and intrachromosomal synteny between these genomes[21,22,24,25] have been hypothesized to be a general feature of such stable bird genomes.[57]

Yet, it is worth noting that the zebra finch lineage underwent high amounts of intrachromosomal rearrangements, such as inversions.[22,128–130] Recent studies involving up to 21 bird genome assemblies suggest additional evidence for frequent interchromosomal and intrachromosomal rearrangements during bird evolution.[23,128,131] Notably, these rearrangements occurred via nonrandom chromosome breakage in frequently reused breakpoint regions,[128,131,132] most of which are associated with elevated densities of TEs.[131,132] It is important to keep in mind that most of the analyzed genomes lack chromosome-scale information and are based on short-read sequencing. Thus, we expect that the actual number of chromosomal rearrangements is much higher, as previously suggested by various cytogenetic studies (e.g., Refs. 127, 129, and 133–136; see Fig. 2 of Ref. 58 for an overview). It is further possible that TEs are even more significantly involved in mediating chromosome breakage than currently anticipated, given that repeat-rich genomic regions remain poorly assembled in second-generation genome assemblies.[111,112]

Even when considering the aforementioned karyotypes of Galliformes and Passeriformes (two-thirds of all bird species) as stable, we argue that at this point it is premature to assume that the same is the case for the remaining one-third of all bird species (33 of 35 avian orders *sensu* Fig. 4). There is cytogenetic evidence for twofold to threefold variation in chromosome numbers within some avian orders (Fig. 4) (e.g., Bucerotiformes, Charadriiformes, Piciformes), some of which coincide with lineages that witnessed increased densities or diversities of TEs (Fig. 4) (e.g., Bucerotiformes, Piciformes). An example among the available genome assemblies is the peregrine falcon, with a diploid karyotype of only 50 chromosomes and cytogenetic evidence for frequent chromosomal fusions.[135] Yet, the lack of a chromosome-level assembly impedes the comparative genomic analysis of how exactly this derived karyotype evolved from ~80 ancestral chromosomes. We thus emphasize the need for *de novo* chromosome-level assemblies (e.g., based on linkage maps[24]) from all avian orders, ideally complemented with cytogenetic data (e.g., reciprocal chromosome painting[133,134]), in order to properly assess which genomes (and genomic regions) of birds are stable. Understanding avian karyotype evolution requires cytogenomics—the integration of the wealth of existing cytogenetic data (e.g., reviewed by Ref. 58) into avian genome assemblies.

The genomic microcosm of transposons and repressors

What exactly are genomes? Rice[137] recently paraphrased Dobzhansky[138] stating that "Nothing in genetics makes sense except in light of genomic conflict." From the perspective of genomic parasites, such as transposons and viruses, genomes of cellular organisms are not simply strings of DNA but complex microcosms of interactions between and among parasitic genes and host genes. It is worth noting that the cellular genome is not a closed system either, given that retroviruses and some dsDNA viruses blur the lines between the transpososphere and the virosphere.[37–39]

Transposon repressors in a nutshell

Nevertheless, much less is known about the transpososphere than about the virosphere and their impacts on cellular organisms. Over the last

years, there has been a growing appreciation of the importance of host–virus interactions,[139] and there are now hundreds of host proteins known to be involved in antiviral response.[140] A recent study conservatively estimated that about a third of all adaptive mutations in human proteins occurred as a response to viruses.[141] It is thus not farfetched to consider viruses as a dominant driver of adaptive evolution of their cellular hosts. On the other hand, our knowledge of transposon repression is still in its infancy. The list of known mechanisms is currently growing exponentially (see Table 1 for a selection and Ref. 142 for an exhaustive review), and we suggest that it is possible that transposons are similarly important drivers of adaptive evolution. Given that TEs are intragenomic endoparasites, the host's defense against TEs can occur on many different levels (Table 1)—before TE transcription (DNA methylation, histone modifications, DNA hypermutation), during TE transcription (premature termination), after TE transcription (RNA hypermutation, piRNA binding, siRNA binding), and before TE transposition (inhibition of ribonucleoprotein complexes).

However, despite the wealth of available bird genomes, studies on bird–TE interactions have been limited to only two TE repression mechanisms. First, a recent study of APOBEC cytosine deaminases across 123 vertebrates reported the vertebrate-wide strongest signals for C-to-U hypermutation of LTR retrotransposons in birds, namely oscine passerines (especially zebra finch and medium ground finch) followed by various nonpasserine Neoaves (kea, white-tailed tropicbird, Northern fulmar, carmine bee-eater, and Adelie penguin).[143] This suggests that birds are very efficient at interfering with LTR reverse transcription, thereby hypermutating novel LTR insertions and decreasing their potential for mobility. Second, cytosine methylation of CpG and non-CpG sites via DNA methyltransferases is a widely studied mechanism for silencing TEs and regulating host gene expression (reviewed in Refs. 42, 144, and 145). In birds, DNA methylation has been studied in chickens[146,147] and the great tit,[25,148] yet three of these studies focused only on gene body and promoter methylation. The only study on TE methylation found that non-CpG methylation explains most of the TE silencing in the great tit.[148] However, the lack of a *de novo* repeat annotation precluded the analysis of TEs specific to the great tit lineage, and it remains to be seen how recently inserted intact TEs are methylated in bird genomes. There are now promising algorithms for confidently identifying methylation states of individual TEs,[149] and we foresee that such analyses are soon feasible for birds with third-generation genome assemblies and in-depth TE annotations.

Bird–TE coevolution

A growing appreciation of bird–TE arms races will have important consequences for the popular field of avian speciation and population genomics.[14–18] In addition to their aforementioned roles in mediating chromosomal rearrangements[131] and other large-scale structural variation,[97,115] individual TE insertions can have many different consequences for the transcriptional regulation of adjacent genomic regions (e.g., reviewed by Refs. 29 and 150). For example, intact internal promoters of a TE may increase proximal gene expression, and, conversely, DNA methylation of a TE may decrease proximal gene expression. The latter can arise from DNA methylation effects of a new TE insertion on nearby CpG sites, a phenomenon that has been called "sloping shores."[151] Assuming that such effects may also occur in presence/absence polymorphisms of very recently inserted TEs, it is possible that they are responsible for some of the patterns of differential methylation and gene expression identified in birds (e.g. Refs. 146 and 152). Furthermore, the aforementioned TE–repressor arms races may affect bird speciation directly. Owing to continuous bird–TE arms races, geographically isolated populations or young species might quickly diverge in their genomic microcosm of TE repressor systems. Secondary contact via hybridization can unlink a TE and its repressor during meiosis of the F1 generation (Fig. 5 of Ref. 153). A subsequent TE burst due to TE de-repression in F1 gametes or F2 offspring can be expected to reduce hybrid fitness and maintain reproductive isolation.[153]

In this context, we note that abundance and diversity of LTR retrotransposons in oscine passerines coincide with expansions of immune gene families, namely the major histocompatibility complex,[154,155] defensins,[156] and Toll-like receptor 7.[157] It remains unknown whether these gene family expansions pre- or postdated LTR diversification, and how exactly they interact with the lineage-specific LTR

Table 1. A selection of recently discovered transposon repression systems in animals

Name	Protein/RNA family	Function	Study organism	Reference
ADARs	Adenosine deaminases acting on RNA	A-to-I hypermutation of TE RNA	Fruit fly	Savva *et al.*[191]
APOBECs	Cytidine deaminases	C-to-U hypermutation of TE DNA	123 vertebrates including 48 birds	Knisbacher & Levanon[143]
CAF-1	Histone chaperone chromatin assembly factor 1	Histone modifications of TE DNA	Mouse	Hatanaka *et al.*[192]
DNMTs	DNA methyltransferases	Cytosine methylation of TE DNA	Human, great tit, many other eukaryotes	Bourc'his & Bestor,[193] Derks *et al.*,[148] many other studies (e.g., Ref. 145)
HENMT1	HEN methyltransferase 1	piRNA methylation	Mouse	Lim *et al.*[194]
IFIT1	Interferon-induced with tetratricopeptide repeats protein	Inhibition of TEs lacking 2′*O*-methylation of mRNA 5′ caps	Human, mouse	Daugherty *et al.*[195]
MORC1	Microrchidia family of GHKL ATPases	Mediation of TE DNA methylation	Mouse	Pastor *et al.*[196]
MOV10	RNA helicase	Binding to TE ribonucleoprotein particle	Human	Goodier *et al.*[197]
Piwi/Asterix/ Panoramix	Piwi protein complex	piRNA silencing	Fruit fly	Yu *et al.*[198]
Piwi/Aub/Ago	Piwi protein complex	piRNA production	Mouse, fruit fly	Han *et al.*[199], Wang *et al.*[200]
SAMHD1	Aicardi-Goutières syndrome gene product	Inhibition of retrotransposition	Human	Zhao *et al.*[201]
Setdb1	H3K9me3 histone methyltransferase	Histone modifications of TE DNA	Mouse	Pasquarella *et al.*[202]
Sumo2 and Trim28	Sumoylation factor 2 and tripartite motif–containing 28 chromatin modifier	Histone modifications of retrovirus DNA	Human	Yang *et al.*[203]
Trim33	Tripartite motif–containing 33 chromatin modifier	Histone modifications of retrovirus DNA	Mouse	Isbel *et al.*[204]
ZAP	Zinc-finger antiviral protein	Inhibition of retrotransposition	Human, mouse, zebrafish	Moldovan & Moran[205]
ZNF91/93	KRAB zinc-finger genes 91/93	Guidance of Trim28 for histone modifications of TE DNA	Primates	Jacobs *et al.*[159]

Note: To date, bird genomes have only been analyzed with reference to APOBECs (C-to-U hypermutation of TEs)[143] and DNMTs (cytosine methylation of TEs).[148]

retrotransposons. We thus predict a bright future for efforts illuminating the coevolution of immune gene families and LTR retrotransposons. We hypothesize that the striking diversity of LTR retrotransposons in oscine passerines influenced their diversification into thousands of species.

Finally, under certain circumstances or genomic contexts, TEs may have adaptively advantageous

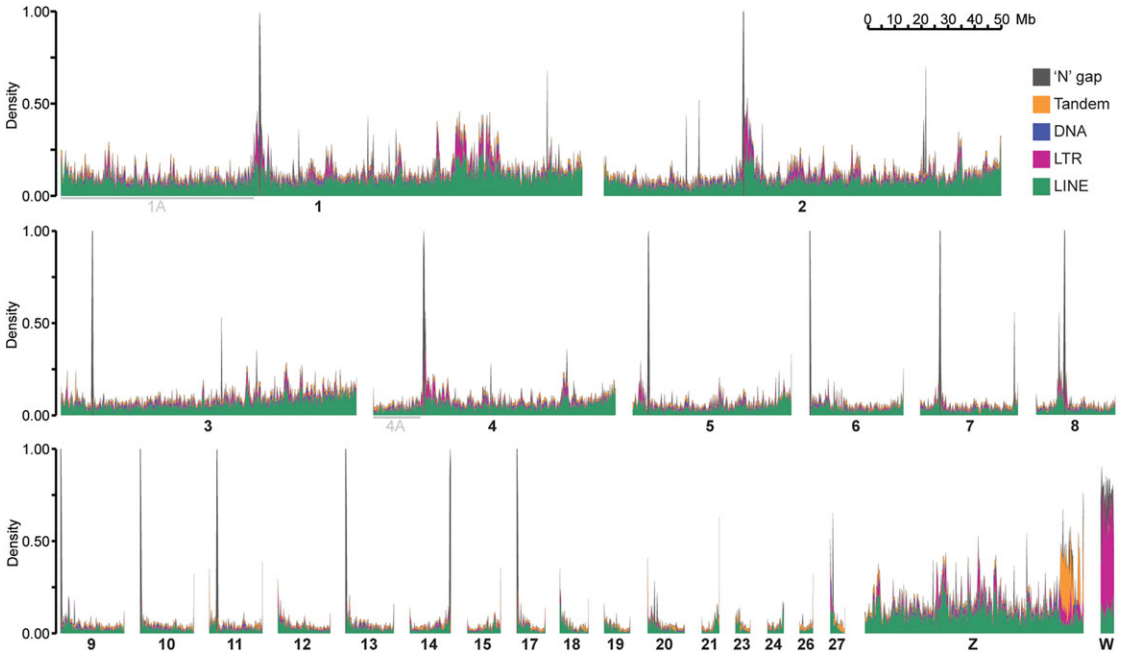

Figure 5. The chromosomal landscape of transposable elements in the third-generation chicken genome. The galGal5 assembly is the first gold-quality genome assembly of a bird, achieved by gap filling of the previous assembly galGal4 via PacBio long-read sequencing (see Ref. 206). It is the only bird genome assembly with known centromere positions on most chromosomes and has the highest sequence contiguity to date (2.9 Mb contig N50; Table S1, online only). Shown are repeat densities for all chromosomes with an assembly size >5 Mb. Repeats were annotated using RepeatMasker[162] with the "-species Aves" option. Densities of TEs, assembly gaps, and tandem repeats (low-complexity repeats, simple repeats, and satellites) were calculated in windows of 200 kb. The plots were generated in R (Supplementary Data S1, online only). Gray bars and gray chromosome numbers indicate karyotype differences between chicken and zebra finch because of fusion/fission events.[22]

effects on the host or may even be coopted as novel genes (reviewed in Refs. 29 and 150). For example, OVEX1, the only coopted TE described in birds, appears to be involved in ovarian differentiation in chicken.[158] But irrespective of whether individual TE insertions interact with the host as parasites, commensals, or symbionts, every "genome is itself in an internal arms race with its own DNA, and thereby inexorably driven toward greater complexity."[159] We anticipate that close inspection of bird genomes will reveal a smörgåsbord of host–TE interactions with multifaceted consequences for bird evolution.

Third-generation genome assemblies and the future of avian genomics

Where will avian genomics research go from here? As mentioned above, the limitations of current second-generation genome assemblies are that tandem repeats and recent TE insertions are underrepresented, especially in structurally complex and repeat-rich regions, and thus lead to assembly gaps and scaffold ends.[111,112] The only way to overcome these issues is to utilize read lengths longer than individual tandem repeat units and full-length TEs. This can be achieved via long-read sequencing of single molecules (reviewed in Refs. 111, 112). The recent generation of a PacBio-based assembly of the TE-rich gorilla genome successfully demonstrated sequence resolution of even the largest TEs,[160] many of which previously constituted assembly gaps due to their size or between-copy sequence similarity (e.g., see Fig. 5 and Table S11 of Ref. 160). Ongoing efforts to generate such third-generation genome assemblies for various birds all yield highly contiguous, gap-free sequences (i.e., contig N50 >>1 Mb; Erich D. Jarvis, personal communication). At present, a PacBio-updated chicken genome assembly is available in GenBank (assembly version galGal5; see Ref. 206). Below, we will discuss the

consequences of such a third-generation genome assembly for the upcoming years of avian genomics research.

Why transposon annotation matters

With major improvements in sequence contiguity in upcoming third-generation genome assemblies due to the resolution of structurally complex and repeat-rich regions, accurate annotation of TEs will be more important than ever. However, this was already evident from analysis of second-generation genome assemblies. A recent in-depth study[161] suggests that repeat annotation does not simply entail the RepeatMasker[162] implementation of repeat libraries available from Repbase[38,163] or automated *de novo* predictions (e.g., RepeatModeler[164]), as these may significantly underestimate TE densities and overestimate TE ages (see Fig. 2 of Ref. 161). Furthermore, many TEs are often misclassified or labeled "unknown" in automated classifications. The only way to obviate this is via manual inspection of alignments of multiple copies from each TE family, followed by generation of curated TE consensus sequences[161] or profile hidden Markov models.[165] We note that manual curation is particularly important for LTR retrotransposons, because their solo-LTRs evolve so rapidly that most are automatically classified as unknown repeats. Manual curation of automated *de novo* predictions in collared flycatcher[102] and hooded crow[103] revealed many lineage-specific LTR retrotransposons with no sequence similarity to previously known solo-LTRs (A. Suh, work in progress).

Finally, even genomes with existing in-depth annotations are up for surprises. The previous second-generation genome assemblies of chicken (versions galGal3 and galGal4) were already annotated via manually curated TE consensus sequences,[20,83] but two recent analyses (published back to back) each increased the annotated TE content of the chicken genome. First, a targeted search for LTR retrotransposons with the newly developed LocaTR tool[166] revealed an additional ∼14 Mb, increasing the amount of annotated TEs by 14% to a total of ∼113 Mb (i.e., a total TE content of ∼11%).[166] Second, the re-annotation of repetitive elements of the galGal4 assembly with an independent set of *de novo* tools well benchmarked for plants (REPET[167]) yielded a total TE content of ∼16%.[168] Together, these data suggest that TE content (especially that of recently active TEs[166,168]) is generally underestimated in genome assemblies and requires further in-depth attention in other (bird) genomes.

Repetitive landscape of a third-generation genome assembly

The upcoming galGal5 chicken genome assembly is the first GQ assembly of a bird (Ref. 206 and Table S1, online only). In comparison with galGal4, the galGal5 total repeat copy numbers and total repeat base pairs (annotated with RepeatMasker using the same repeat library) increased significantly from ∼131 to ∼202 Mb, which implies that >35% of avian repetitive DNA is missing from current second-generation genome assemblies. Among TEs, this includes ∼11 Mb of newly assembled CR1 elements alone. Most strikingly, the long PacBio reads permitted the resolution of many large tandem repeats, such as ∼54 Mb of satellite repeats, which previously resisted assembly. Third-generation genome assembly exhibits improved sequence resolution of the microchromosomes and an overall high contiguity, with the promise of resolving large TEs in place of assembly gaps.

Such highly contiguous chromosome-level genome assembly allows for new observations of the chromosomal distribution of TEs, tandem repeats, and assembly gaps (Fig. 5; windows with 100% gaps are centromere placeholders). Although the gene-rich microchromosomes generally exhibit lower repeat densities than the gene-poor macrochromosomes (also noted in, e.g., Refs. 48, 59, and 169), there is considerable within-chromosome variation in the distribution of these genomic features (Fig. 5). There are many chromosomal regions with local TE densities of ∼50% in 200-kb windows, especially on large macrochromosomes, the Z chromosome, and near centromeres. Notably, this high TE density is comparable to the situation in the TE-rich mammalian genomes.[90,160] Even more repetitive is the W chromosome, which will be discussed below.

Many of these TE-rich regions previously resisted assembly for the aforementioned technological reasons. We thus hypothesize that a genomic division into TE-poor (gene-rich) and TE-rich (gene-poor) regions is a common feature of bird genomes. This pattern is reminiscent of the two-speed genome concept (originally introduced for plant-pathogenic fungi[170,171]) which suggests evolutionary stability

in gene-rich regions and evolutionary instability in TE-rich regions. Not surprisingly, TEs have already been implicated as playing an important role in break point reuse for avian chromosomal rearrangements.[131,132] The commonly acknowledged stability of bird genomes[57] may thus apply to many, but by far not all, parts of the genome.

Highly repeat-rich regions

Even in the GQ chicken assembly (galGal5), some chromosomes are either entirely missing (chromosomes 34–38) or poorly assembled owing to their putative repetitiveness (e.g., chromosomes 16, 22, and 25[172]). Strikingly, the TE-richest assembled chromosome appears to be the female-specific W chromosome (∼ 68% TEs (Fig. 5)). Rapid accumulation of TEs and tandem repeats on the nonrecombining W (reviewed in Refs. 173 and 174) has been implicated in suppression of recombination during Z–W differentiation.[175] Even Illumina-based W drafts are TE rich (∼50% in flycatcher[102] and white-throated sparrow[176]), and we hypothesize that avian W chromosomes are a TE "refugium," especially for LTR retrotransposons.

Within chromosomes, the TE-richest regions appear to be near centromeres (Fig. 5). Centromeres themselves consist of long tandem arrays of satellite repeats bound by the spindle apparatus during cell division (reviewed in Ref. 177). For birds, direct evidence for centromere locations is, however, limited to the chicken, where there are gap placeholders on 14 chromosomes (Fig. 5). Interestingly, chicken chromosomes 5, 27, and Z may contain short, nontandem-repetitive centromeres,[178] and some of the satellite repeats of macrochromosome centromeres contain CR1-derived subunits.[178,179] A body of nonavian evidence suggests direct involvement of TEs in shaping centromere architecture (reviewed in Refs. 180 and 181), and we note that centromeric TEs can thereby affect speciation through the meiotic process of centromere drive (reviewed in Refs. 182 and 183). Thus, the impact of centromeres on bird evolution warrants further study despite current methodological limitations of assembling 100 kb–scale tandem repeat arrays.[184]

These setbacks aside, we anticipate a rapid onset of published third-generation genome assemblies of birds that will spur avian genomics research on TE-rich structurally complex regions. The plethora of novel methods combining multiplatform data into chromosome-level assemblies (e.g., Refs. 185 and 186) will further facilitate these developments.

Conclusions

On the quest toward complete genomes (if they can ever be achieved), genomics research (not just of birds) is currently undergoing a major transition from draft short-read genome assemblies to highly contiguous long-read genome assemblies. Simultaneously, TEs, EVEs, and other repetitive elements are becoming widely appreciated in the context of assembly, annotation, and evolution of genomes. In particular, TEs can have important consequences for genomic instability, transcriptional regulation, and reproductive isolation of their hosts. We propose that birds exhibit genomes with both stable gene-rich and unstable TE-rich regions. The latter are now finally accessible because of the power of third-generation genome assemblies. We predict that integrated cytogenomics of structurally complex regions will reveal additional hidden dynamics of the small and streamlined bird genomes.

Acknowledgments

We thank David W. Burt, Daren C. Card, Tyler A. Elliott, Lel Eory, Cédric Feschotte, Carolina Frankl-Vilches, Dustin C. Hancks, Erich D. Jarvis, Ulrich Knief, Kenji K. Kojima, Heiner Kuhl, Cai Li, Andrew S. Mason, Kees van Oers, Chris Organ, Homa Papoli Yazdi, Gabriel L. Wallau, Wesley C. Warren, and Matthias Weissensteiner for helpful discussions and unpublished information. We further thank Reto Burri, Homa Papoli Yazdi, and two anonymous reviewers for comments on earlier versions of this manuscript. A.S. acknowledges a Junior Researcher Grant of the SciLifeLab Swedish Biodiversity Program to generate third-generation genome assemblies from birds-of-paradise. We thank Jon Fjeldså for providing the bird paintings.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1. Avian genome assemblies publicly available as of June 2016.

Data S1. R script to plot the chromosome landscapes of Figure 5.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Xu, X., Z. Zhou, R. Dudley, *et al.* 2014. An integrative approach to understanding bird origins. *Science* **346:** 1253293.
2. Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection.* London: John Murray.
3. Holt, B.G., J.-P. Lessard, M.K. Borregaard, *et al.* 2013. An update of Wallace's zoogeographic regions of the world. *Science* **339:** 74–78.
4. Jetz, W., G. Thomas, J. Joy, *et al.* 2012. The global diversity of birds in space and time. *Nature* **491:** 444–448.
5. Shapiro, M.D., Z. Kronenberg, C. Li, *et al.* 2013. Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339:** 1063–1067.
6. Towers, M., J. Signolet, A. Sherman, *et al.* 2011. Insights into bird wing evolution and digit specification from polarizing region fate maps. *Nat. Commun.* **2:** 426.
7. Rubin, C.-J., M.C. Zody, J. Eriksson, *et al.* 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464:** 587–591.
8. Jarvis, E.D., S. Ribeiro, M.L. da Silva, *et al.* 2000. Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature* **406:** 628–632.
9. Fisher, S.E. & C. Scharff. 2009. FOXP2 as a molecular window into speech and language. *Trends Genet.* **25:** 166–177.
10. Jarvis, E.D., S. Mirarab, A.J. Aberer, *et al.* 2014. Whole genome analyses resolve the early branches in the tree of life of modern birds. *Science* **346:** 1320–1331.
11. Prum, R.O., J.S. Berv, A. Dornburg, *et al.* 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526:** 569–573.
12. Suh, A., L. Smeds & H. Ellegren. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol.* **13:** e1002224.
13. Suh, A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zool. Scripta* **45:** 50–62.
14. Lamichhaney, S., J. Berglund, M.S. Almen, *et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518:** 371–375.
15. Burri, R., A. Nater, T. Kawakami, *et al.* 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum in *Ficedula* flycatchers. *Genome Res.* **25:** 1656–1665.
16. Poelstra, J.W., N. Vijay, C.M. Bossu, *et al.* 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344:** 1410–1414.
17. Ellegren, H., L. Smeds, R. Burri, *et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491:** 756–760.
18. Burri, R., S. Antoniazza, A. Gaigher, *et al.* 2016. The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution* **70:** 140–153.
19. Koepfli, K.-P., B. Paten & S.J. O'Brien. 2015. The Genome 10K Project: a way forward. *Annu. Rev. Anim. Biosci.* **3:** 57–111.
20. Hillier, L.W., W. Miller, E. Birney, *et al.* 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695–716.
21. Dalloul, R.A., J.A. Long, A.V. Zimin, *et al.* 2010. Multiplatform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8:** 1–21.
22. Warren, W.C., D.F. Clayton, H. Ellegren, *et al.* 2010. The genome of a songbird. *Nature* **464:** 757–762.
23. Frankl-Vilches, C., H. Kuhl, M. Werber, *et al.* 2015. Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. *Genome Biol.* **16:** 19.
24. Kawakami, T., L. Smeds, N. Backström, *et al.* 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* **23:** 4035–4058.
25. Laine, V.N., T.I. Gossmann, K.M. Schachtschneider, *et al.* 2016. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **7:** 10474.
26. Zhang, G., C. Li, Q. Li, *et al.* 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346:** 1311–1320.
27. Zhang, J., C. Li, Q. Zhou, *et al.* 2015. Improving the ostrich genome assembly using optical mapping data. *Gigascience* **4:** 24.
28. Jarvis, E.D. 2016. Perspectives from the Avian Phylogenomics Project: questions that can be answered with sequencing all genomes of a vertebrate class. *Annu. Rev. Anim. Biosci.* **4:** 45–59.
29. Kazazian, H.H, Jr. 2004. Mobile elements: drivers of genome evolution. *Science* **303:** 1626–1632.
30. Kidwell, M. 2005. Transposable elements. In *The Evolution of the Genome.* T.R. Gregory, Ed.: 165–221. Elsevier Academic Press.
31. Katzourakis, A. & R.J. Gifford. 2010. Endogenous viral elements in animal genomes. *PLoS Genet.* **6:** e1001191.
32. Suh, A., C.C. Weber, C. Kehlmaier, *et al.* 2014. Early Mesozoic coexistence of amniotes and Hepadnaviridae. *PLoS Genet.* **10:** e1004559.
33. Aiewsakun, P. & A. Katzourakis. 2015. Endogenous viruses: connecting recent and ancient viral evolution. *Virology* **479–480:** 26–37.
34. Holmes, E.C. 2011. The evolution of endogenous viral elements. *Cell Host Microbe* **10:** 368–377.
35. Feschotte, C. & C. Gilbert. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13:** 283–296.
36. Weiss, R.A. & J.P. Stoye. 2013. Our viral inheritance. *Science* **340:** 820–821.

37. Wicker, T., F. Sabot, A. Hua-Van, *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8:** 973–982.
38. Kapitonov, V.V. & J. Jurka. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **9:** 411–412.
39. Krupovic, M. & E.V. Koonin. 2015. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol.* **13:** 105–115.
40. Lee, Y.C.G. & C.H. Langley. 2010. Transposable elements in natural populations of *Drosophila melanogaster*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365:** 1219–1228.
41. Elliott, T.A. & T.R. Gregory. 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:** 20140331. doi: 10.1098/rstb.2014.0331.
42. Slotkin, R.K. & R. Martienssen. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8:** 272–285.
43. Lowe, C.B., G. Bejerano & D. Haussler. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc. Natl. Acad. Sci. U.S.A.* **104:** 8005–8010.
44. Schmidt, D., P.C. Schwalie, M.D. Wilson, *et al.* 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148:** 335–348.
45. Wang, J., C. Vicente-García, D. Seruggia, *et al.* 2015. MIR retrotransposon sequences provide insulators to the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **112:** E4428–E4437.
46. Sotero-Caio, C., R.N. Platt II, A. Suh, *et al.* 2016. Evolution and diversity of transposable elements in vertebrates. *Genome Biol. Evol.* doi: 10.1093/gbe/evw264.
47. Gregory, T.R. 2016. Animal genome size database. http://www.genomesize.com. Last accessed May 27, 2016.
48. Organ, C.L. & S.V. Edwards. 2011. Major events in avian genome evolution. In *Living Dinosaurs: The Evolutionary History of Modern Birds.* G. Dyke & G. Kaiser, Eds.: 325–337. John Wiley & Sons, Ltd.
49. Gregory, T.R., C.B. Andrews, J.A. McGuire, *et al.* 2009. The smallest avian genomes are found in hummingbirds. *Proc. Biol. Sci.* **276:** 3753–3757.
50. Organ, C.L., A.M. Shedlock, A. Meade, *et al.* 2007. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446:** 180–184.
51. Wright, N.A., T.R. Gregory & C.C. Witt. 2014. Metabolic 'engines' of flight drive genome size reduction in birds. *Proc. Biol. Sci.* **281:** 20132780.
52. Zhang, Q. & S.V. Edwards. 2012. The evolution of intron size in amniotes: a role for powered flight? *Genome Biol. Evol.* **4:** 1033–1043.
53. Hughes, A.L. & M.K. Hughes. 1995. Small genomes for better flyers. *Nature* **377:** 391.
54. Vinogradov, A.E. 1997. Nucleotypic effect in homeotherms: body-mass independent resting metabolic rate of passerine birds is related to genome size. *Evolution* **51:** 220–225.
55. Schubert, I. & G.T.H. Vu. 2016. Genome stability and evolution: attempting a holistic view. *Trends Plant Sci.* **21:** 749–757.
56. Oliver, M.J., D. Petrov, D. Ackerly, *et al.* 2007. The mode and tempo of genome size evolution in eukaryotes. *Genome Res.* **17:** 594–601.
57. Ellegren, H. 2010. Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* **25:** 283–291.
58. Deakin, J.E. & T. Ezaz. 2014. Tracing the evolution of amniote chromosomes. *Chromosoma* **123:** 201–216.
59. Burt, D.W. 2002. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96:** 97–112.
60. Organ, C.L., R.G. Moreno & S.V. Edwards. 2008. Three tiers of genome evolution in reptiles. *Integr. Comp. Biol.* **48:** 494–504.
61. Ellegren, H. & N. Galtier. 2016. Determinants of genetic diversity. *Nat. Rev. Genet.* **17:** 422–433.
62. Kordiš, D. 2009. Transposable elements in reptilian and avian (Sauropsida) genomes. *Cytogenet. Genome Res.* **127:** 94–111.
63. Chalopin, D., M. Naville, F. Plard, *et al.* 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **7:** 567–580.
64. Janes, D.E., C.L. Organ, M.K. Fujita, *et al.* 2010. Genome evolution in Reptilia, the sister group of mammals. *Annu. Rev. Genomics Hum. Genet.* **11:** 239–264.
65. Cui, J., W. Zhao, Z. Huang, *et al.* 2014. Low frequency of paleoviral infiltration across the avian phylogeny. *Genome Biol.* **15:** 539.
66. Suh, A., G. Churakov, M.P. Ramakodi, *et al.* 2015. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol. Evol.* **7:** 205–217.
67. Shedlock, A.M., C.W. Botka, S. Zhao, *et al.* 2007. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl. Acad. Sci. U.S.A.* **104:** 2767–2772.
68. Suh, A. 2015. The specific requirements for CR1 retrotransposition explain the scarcity of retrogenes in birds. *J. Mol. Evol.* **81:** 18–20.
69. Ivancevic, A.M., R.D. Kortschak, T. Bertozzi, *et al.* 2016. LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. *Genome Biol. Evol.* doi: 10.1093/gbe/evw243.
70. Ichiyanagi, K. & N. Okada. 2008. Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. *Mol. Biol. Evol.* **25:** 1148–1157.
71. Luan, D.D., M.H. Korman, J.L. Jakubczak, *et al.* 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72:** 595–605.
72. Martin, S.L. 2006. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J. Biomed. Biotechnol.* **2006:** 6.
73. Levin, H.L. & J.V. Moran. 2011. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12:** 615–627.
74. Stumph, W.E., P. Kristo, M.-J. Tsai, *et al.* 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. *Nucleic Acids Res.* **9:** 5383–5398.

75. Stumph, W.E., C.P. Hodgson, M.J. Tsai, *et al.* 1984. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. *Proc. Natl. Acad. Sci. U.S.A.* **81:** 6667–6671.
76. Lovšin, N., F. Gubenšek & D. Kordiš. 2001. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. *Mol. Biol. Evol.* **18:** 2213–2224.
77. Vandergon, T.L. & M. Reitman. 1994. Evolution of chicken repeat 1 (CR1) elements: evidence for ancient subfamilies and multiple progenitors. *Mol. Biol. Evol.* **11:** 886–898.
78. Kriegs, J.O., A. Matzke, G. Churakov, *et al.* 2007. Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol. Biol.* **7:** 190.
79. Suh, A., M. Paus, M. Kiefmann, *et al.* 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.* **2:** 443.
80. Suh, A., J.O. Kriegs, S. Donnellan, *et al.* 2012. A universal method for the study of CR1 retroposons in nonmodel bird genomes. *Mol. Biol. Evol.* **29:** 2899–2903.
81. Matzke, A., G. Churakov, P. Berkes, *et al.* 2012. Retroposon insertion patterns of neoavian birds: strong evidence for an extensive incomplete lineage sorting era. *Mol. Biol. Evol.* **29:** 1497–1501.
82. Kaiser, V.B., M. van Tuinen & H. Ellegren. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol. Biol. Evol.* **24:** 338–347.
83. Wicker, T., J.S. Robertson, S.R. Schulze, *et al.* 2005. The repetitive landscape of the chicken genome. *Genome Res.* **15:** 126–136.
84. Suh, A., C.C. Witt, J. Menger, *et al.* 2016. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *Nat. Commun.* **7:** 11396.
85. Kordiš, D. & F. Gubenšek. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc. Natl. Acad. Sci. U.S.A.* **95:** 10704–10709.
86. Walsh, A.M., R.D. Kortschak, M.G. Gardner, *et al.* 2013. Widespread horizontal transfer of retrotransposons. *Proc. Natl. Acad. Sci. U.S.A.* **110:** 1012–1016.
87. Kojima, K.K. & H. Fujiwara. 2005. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol. Biol. Evol.* **22:** 2157–2165.
88. Kojima, K.K., Y. Seto & H. Fujiwara. 2016. The wide distribution and change of target specificity of R2 non-LTR retrotransposons in animals. *PLoS One* **11:** e0163496.
89. Ohshima, K., M. Hamada, Y. Terai, *et al.* 1996. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. *Mol. Cell. Biol.* **16:** 3756–3764.
90. Lander, E.S., L.M. Linton, B. Birren, *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
91. Hirakawa, M., H. Nishihara, M. Kanehisa, *et al.* 2009. Characterization and evolutionary landscape of AmnSINE1 in Amniota genomes. *Gene* **441:** 100–110.
92. Bejerano, G., C.B. Lowe, N. Ahituv, *et al.* 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441:** 87–90.
93. Nishihara, H., A.F.A. Smit & N. Okada. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* **16:** 864–874.
94. Green, R.E., E.L. Braun, J. Armstrong, *et al.* 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346:** 1335.
95. Suh, A., S. Bachg, S. Donnellan, *et al.* 2016. *De-novo* emergence and template switching of SINE retroposons during the early evolution of passerine birds. *bioRxiv* doi: 10.1101/081950.
96. Gogolevsky, K.P., N.S. Vassetzky & D.A. Kramerov. 2008. Bov-B-mobilized SINEs in vertebrate genomes. *Gene* **407:** 75–85.
97. Devos, K.M., J.K.M. Brown & J.L. Bennetzen. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12:** 1075–1079.
98. Kovalskaya, E., A. Buzdin, E. Gogvadze, *et al.* 2006. Functional human endogenous retroviral LTR transcription start sites are located between the R and U5 regions. *Virology* **346:** 373–378.
99. Hayward, A., C.K. Cornwallis & P. Jern. 2015. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc. Natl. Acad. Sci. U.S.A.* **112:** 464–469.
100. Henzy, J.E., R.J. Gifford, W.E. Johnson, *et al.* 2014. A novel recombinant retrovirus in the genomes of modern birds combines features of avian and mammalian retroviruses. *J. Virol.* **88:** 2398–2405.
101. Suh, A., J.O. Kriegs, J. Brosius, *et al.* 2011. Retroposon insertions and the chronology of avian sex chromosome evolution. *Mol. Biol. Evol.* **28:** 2993–2997.
102. Smeds, L., V. Warmuth, P. Bolivar, *et al.* 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nat. Commun.* **6:** 7330.
103. Vijay, N., C.M. Bossu, J.W. Poelstra, *et al.* 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Commun.* **7:** 13195.
104. Feschotte, C. & E.J. Pritham. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41:** 331–368.
105. Suh, A., J. Brosius, J. Schmitz, *et al.* 2013. The genome of a Mesozoic paleovirus reveals the evolution of hepatitis B viruses. *Nat. Commun.* **4:** 1791.
106. Cui, J. & E.C. Holmes. 2012. Endogenous hepadnaviruses in the genome of the budgerigar (*Melopsittacus undulatus*) and the evolution of avian hepadnaviruses. *J. Virol.* **86:** 7688–7691.
107. Liu, W., S. Pan, H. Yang, *et al.* 2012. The first full-length endogenous hepadnaviruses: identification and analysis. *J. Virol.* **86:** 9510–9513.
108. Gilbert, C. & C. Feschotte. 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. *PLoS Biol.* **8:** e1000495.
109. Orito, E., M. Mizokami, Y. Ina, *et al.* 1989. Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. *Proc. Natl. Acad. Sci. U.S.A.* **86:** 7059–7062.
110. van Hemert, F.J., M.A.A. van de Klundert, V.V. Lukashov, *et al.* 2011. Protein X of hepatitis B virus: origin and

structure similarity with the central domain of DNA glycosylase. *PLoS One* **6:** e23392.
111. Goodwin, S., J.D. McPherson & W.R. McCombie. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17:** 333–351.
112. Chaisson, M.J.P., R.K. Wilson & E.E. Eichler. 2015. Genetic variation and the *de novo* assembly of human genomes. *Nat. Rev. Genet.* **16:** 627–640.
113. Kapusta, A., A. Suh & C. Feschotte. 2016. The hidden elasticity of avian and mammalian genomes. *bioRxiv* doi: 10.1101/081307.
114. Weber, C.C., B. Nabholz, J. Romiguier, *et al.* 2014. K_r/K_c but not d_N/d_S correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol.* **15:** 1–13.
115. Konkel, M.K. & M.A. Batzer. 2010. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. *Semin. Cancer Biol.* **20:** 211–221.
116. van de Lagemaat, L.N., L. Gagnier, P. Medstrand, *et al.* 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res.* **15:** 1243–1249.
117. Klein, J., H. Ono, D. Klein, *et al.* 1993. The accordion model of *Mhc* evolution. In *Progress in Immunology*. Vol. VIII. J. Gergely, M. Benczúr, A. Erdei, Eds.: 137–143. Berlin, Heidelberg: Springer.
118. Lynch, M. & J.S. Conery. 2003. The origins of genome complexity. *Science* **302:** 1401–1404.
119. Lynch, M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U.S.A.* **104:** 8597–8604.
120. Lanfear, R., S.Y.W. Ho, D. Love, *et al.* 2010. Mutation rate is linked to diversification in birds. *Proc. Natl. Acad. Sci. U.S.A.* **107:** 20423–20428.
121. Yandell, M. & D. Ence. 2012. A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **13:** 329–342.
122. Ruiz-Herrera, A., M. Farré & T.J. Robinson. 2012. Molecular cytogenetic and genomic insights into chromosomal evolution. *Heredity* **108:** 28–36.
123. Belterman, R.H.R. & L.E.M. De Boer. 1990. A miscellaneous collection of bird karyotypes. *Genetica* **83:** 17–29.
124. Belterman, R.H.R. & L.E.M. De Boer. 1984. A karyological study of 55 species of birds, including karyotypes of 39 species new to cytology. *Genetica* **65:** 39–82.
125. Bian, X., H. Cai, S. Ning, *et al.* 1991. Studies on the karyotypes of birds XII. 15 species of nonpasserines (Aves). *Zool. Res.* **4:** 016.
126. Garnero, A.D.V., M. Boccelli, J.C.P. Oliveira, *et al.* 2013. Chromosomal characterization of four Antarctic Procellariiformes. *Mar. Ornithol.* **41:** 63–68.
127. Griffin, D.K., L.B.W. Robertson, H.G. Tempest, *et al.* 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet. Genome Res.* **117:** 64–77.
128. Romanov, M., M. Farre, P. Lithgow, *et al.* 2014. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC Genomics* **15:** 1060.
129. Hooper, D.M. & T.D. Price. 2015. Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. *Evolution* **69:** 890–903.
130. Knief, U., G. Hemmrich-Stanisak, M. Wittig, *et al.* 2016. Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biol.* **17:** 1–22.
131. Farré, M., J. Narayan, G.T. Slavov, *et al.* 2016. Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome Biol. Evol.* **8:** 2442–2451.
132. Skinner, B.M. & D.K. Griffin. 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* **108:** 37–41.
133. Nie, W., P.C.M. O'Brien, B.L. Ng, *et al.* 2009. Avian comparative genomics: reciprocal chromosome painting between domestic chicken (*Gallus gallus*) and the stone curlew (*Burhinus oedicnemus*, Charadriiformes)—an atypical species with low diploid number. *Chromosome Res.* **17:** 99–113.
134. Nie, W., P. O'Brien, B. Fu, *et al.* 2015. Multidirectional chromosome painting substantiates the occurrence of extensive genomic reshuffling within Accipitriformes. *BMC Evol. Biol.* **15:** 205.
135. Nishida, C., J. Ishijima, A. Kosaka, *et al.* 2008. Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation. *Chromosome Res.* **16:** 171–181.
136. de Oliveira Furo, I., R. Kretschmer, P.C. O'Brien, *et al.* 2015. Chromosomal diversity and karyotype evolution in South American macaws (Psittaciformes, Psittacidae). *PLoS One* **10:** e0130157.
137. Rice, W.R. 2013. Nothing in genetics makes sense except in light of genomic conflict. *Annu. Rev. Ecol. Evol. Syst.* **44:** 217–237.
138. Dobzhansky, T. 1973. Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* **35:** 125–129.
139. Guirimand, T., S. Delmotte & V. Navratil. 2015. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res.* **43:** D583–D587.
140. Dyer, M.D., T.M. Murali & B.W. Sobral. 2008. The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.* **4:** e32.
141. Enard, D., L. Cai, C. Gwennap, *et al.* 2016. Viruses are a dominant driver of protein adaptation in mammals. *Elife* **5:** e12469.
142. Goodier, J.L. 2016. Restricting retrotransposons: a review. *Mob. DNA* **7:** 16.
143. Knisbacher, B.A. & E.Y. Levanon. 2015. DNA editing of LTR retrotransposons reveals the impact of APOBECs on vertebrate genomes. *Mol. Biol. Evol.* **33:** 554–567.
144. Suzuki, M.M. & A. Bird. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9:** 465–476.
145. Zemach, A., I.E. McDaniel, P. Silva, *et al.* 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328:** 916–919.
146. Nätt, D., C.-J. Rubin, D. Wright, *et al.* 2012. Heritable genome-wide variation of gene expression and

promoter methylation between wild and domesticated chickens. *BMC Genomics* **13:** 1–12.
147. Mugal, C.F., P.F. Arndt, L. Holm, *et al.* 2015. Evolutionary consequences of DNA methylation on the GC content in vertebrate genomes. *G3* **5:** 441–447.
148. Derks, M.F.L., K.M. Schachtschneider, O. Madsen, *et al.* 2016. Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genomics* **17:** 1–13.
149. Suzuki, Y., J. Korlach, S.W. Turner, *et al.* 2015. AgIn: measuring the landscape of CpG methylation of individual repetitive elements. *Bioinformatics* **32:** 2911–2919.
150. Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9:** 397–405.
151. Grandi, F.C., J.M. Rosser, S.J. Newkirk, *et al.* 2015. Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. *Genome Res.* **25:** 1135–1146.
152. Uebbing, S., A. Künstner, H. Mäkinen, *et al.* 2016. Divergence in gene expression within and between two closely related flycatcher species. *Mol. Ecol.* **25:** 2015–2028.
153. Rogers, R.L. 2015. Chromosomal rearrangements as barriers to genetic homogenization between archaic and modern humans. *Mol. Biol. Evol.* **32:** 3064–3078.
154. Balakrishnan, C.N., R. Ekblom, M. Völker, *et al.* 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol.* **8:** 1–19.
155. Eimes, J.A., S.-I. Lee, A.K. Townsend, *et al.* 2016. Early duplication of a single MHC IIB locus prior to the passerine radiations. *PLoS One* **11:** e0163456.
156. Cheng, Y., M.D. Prickett, W. Gutowska, *et al.* 2015. Evolution of the avian β-defensin and cathelicidin genes. *BMC Evol. Biol.* **15:** 1–17.
157. Grueber, C.E., G.P. Wallis, T.M. King, *et al.* 2012. Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS One* **7:** e45011.
158. Carré-Eusèbe, D., N. Coudouel & S. Magre. 2009. OVEX1, a novel chicken endogenous retrovirus with sex-specific and left–right asymmetrical expression in gonads. *Retrovirology* **6:** 1–24.
159. Jacobs, F.M.J., D. Greenberg, N. Nguyen, *et al.* 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516:** 242–245.
160. Gordon, D., J. Huddleston, M.J.P. Chaisson, *et al.* 2016. Long-read sequence assembly of the gorilla genome. *Science* **352:** aae0344.
161. Platt II, R.N., L. Blanco-Berdugo & D.A. Ray. 2016. Accurate transposable element annotation is vital when analyzing new genome assemblies. *Genome Biol. Evol.* **8:** 403–410.
162. Smit, A., R. Hubley & P. Green. 1996–2010. RepeatMasker Open-3.3.0. http://www.repeatmasker.org. Last accessed May 27, 2016.
163. Bao, W., K. Kojima & O. Kohany. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6:** 11. Last accessed May 27, 2016.
164. Smit, A. & R. Hubley. 2010. RepeatModeler Open-1.0. http://www.repeatmasker.org. Last accessed May 27, 2016.
165. Hubley, R., R.D. Finn, J. Clements, *et al.* 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44:** D81–D89.
166. Mason, A.S., J.E. Fulton, P.M. Hocking, *et al.* 2016. A new look at the LTR retrotransposon content of the chicken genome. *BMC Genomics* **17:** 688.
167. Quesneville, H., C.M. Bergman, O. Andrieu, *et al.* 2005. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.* **1:** e22.
168. Guizard, S., B. Piégu, P. Arensburger, *et al.* 2016. Deep landscape update of dispersed and tandem repeats in the genome model of the red jungle fowl, *Gallus gallus*, using a series of *de novo* investigating tools. *BMC Genomics* **17:** 1–23.
169. Abrusán, G., H.-J. Krambeck, T. Junier, *et al.* 2008. Biased distributions and decay of long interspersed nuclear elements in the chicken genome. *Genetics* **178:** 573–581.
170. Raffaele, S. & S. Kamoun. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Micro.* **10:** 417–430.
171. Faino, L., M.F. Seidl, X. Shi-Kunne, *et al.* 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* **26:** 1091–1100.
172. Masabanda, J.S., D.W. Burt, P.C.M. O'Brien, *et al.* 2004. Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. *Genetics* **166:** 1367–1373.
173. Schartl, M., M. Schmid & I. Nanda. 2016. Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma* **125:** 553–571.
174. Wright, A.E., R. Dean, F. Zimmer, *et al.* 2016. How to make a sex chromosome. *Nat. Commun.* **7:** 12087.
175. Suh, A. 2012. A retroposon-based view on the temporal differentiation of sex chromosomes. *Mob. Genet. Elements* **2:** 158–162.
176. Davis, J.K., P.J. Thomas & J.W. Thomas. 2010. A W-linked palindrome and gene conversion in New World sparrows and blackbirds. *Chromosome Res.* **18:** 543–553.
177. Kursel, L.E. & H.S. Malik. 2016. Centromeres. *Curr. Biol.* **26:** R487–R490.
178. Shang, W.-H., T. Hori, A. Toyoda, *et al.* 2010. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res.* **20:** 1219–1228.
179. Li, J. & F.C. Leung. 2006. A CR1 element is embedded in a novel tandem repeat (*Hin*fI repeat) within the chicken genome. *Genome* **49:** 97–103.
180. Meštrović, N., B. Mravinac, M. Pavlek, *et al.* 2015. Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Res.* **23:** 583–596.
181. Carone, D.M. & R.J. O'Neill. 2010. Marsupial centromeres and telomeres: dynamic chromosome domains. In *Marsupial Genetics and Genomics*. E.J. Deakin, D.P. Waters & A.J. Marshall Graves, Eds.: 55–73. Dordrecht: Springer.
182. Henikoff, S., K. Ahmad & H.S. Malik. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293:** 1098–1102.
183. Lindholm, A.K., K.A. Dyer, R.C. Firman, *et al.* 2016. The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* **31:** 315–326.

184. Miga, K.H. 2015. Completing the human genome: the progress and challenge of satellite DNA assembly. *Chromosome Res.* **23:** 421–426.
185. Vij, S., H. Kuhl, I.S. Kuznetsova, *et al.* 2016. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genet.* **12:** e1005954.
186. Bickhart, D.M., B.D. Rosen, S. Koren, *et al.* 2016. Single-molecule sequencing and conformational capture enable *de novo* mammalian reference genomes. *bioRxiv* doi: 10.1101/064352.
187. Shedlock, A.M. & S.V. Edwards. 2009. Amniotes (Amniota). In *The Timetree of Life.* S.B. Hedges & S. Kumar, Eds.: 375–379. New York: Oxford University Press.
188. Doležel, J., J. Bartoš, H. Voglmayr, *et al.* 2003. Nuclear DNA content and genome size of trout and human. *Cytometry A* **51:** 127–128.
189. Shaffer, H.B., P. Minx, D. Warren, *et al.* 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14:** R28.
190. Alföldi, J., F. Di Palma, M. Grabherr, *et al.* 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477:** 587–591.
191. Savva, Y.A., J.E.C. Jepson, Y.-J. Chang, *et al.* 2013. RNA editing regulates transposon-mediated heterochromatic gene silencing. *Nat. Commun.* **4:** 2745.
192. Hatanaka, Y., K. Inoue, M. Oikawa, *et al.* 2015. Histone chaperone CAF-1 mediates repressive histone modifications to protect preimplantation mouse embryos from endogenous retrotransposons. *Proc. Natl. Acad. Sci. U.S.A.* **112:** 14641–14646.
193. Bourc'his, D. & T.H. Bestor. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431:** 96–99.
194. Lim, S.L., Z.P. Qu, R.D. Kortschak, *et al.* 2015. HENMT1 and piRNA stability are required for adult male germ cell transposon repression and to define the spermatogenic program in the mouse. *PLoS Genet.* **11:** e1005620.
195. Daugherty, M.D., A.M. Schaller, A.P. Geballe, *et al.* 2016. Evolution-guided functional analyses reveal diverse antiviral specificities encoded by IFIT1 genes in mammals. *Elife* **5:** e14228.
196. Pastor, W.A., H. Stroud, K. Nee, *et al.* 2014. MORC1 represses transposable elements in the mouse male germline. *Nat. Commun.* **5:** 6795.
197. Goodier, J.L., L.E. Cheung & H.H. Kazazian, Jr. 2012. MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.* **8:** e1002941.
198. Yu, Y., J. Gu, Y. Jin, *et al.* 2015. Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* **350:** 339–342.
199. Han, B.W., W. Wang, C. Li, *et al.* 2015. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* **348:** 817–821.
200. Wang, W., Bo W. Han, C. Tipping, *et al.* 2015. Slicing and binding by Ago3 or Aub trigger Piwi-bound piRNA production by distinct mechanisms. *Mol. Cell* **59:** 819–830.
201. Zhao, K., J. Du, X. Han, *et al.* 2013. Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi–Goutières syndrome-related SAMHD1. *Cell Rep.* **4:** 1108–1115.
202. Pasquarella, A., A. Ebert, G.P. de Almeida, *et al.* 2016. Retrotransposon derepression leads to activation of the unfolded protein response and apoptosis in pro-B cells. *Development* **143:** 1788–1799.
203. Yang, B.X., C.A. El Farran, H.C. Guo, *et al.* 2015. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163:** 230–245.
204. Isbel, L., R. Srivastava, H. Oey, *et al.* 2015. Trim33 binds and silences a class of young endogenous retroviruses in the mouse testis; a novel component of the arms race between retrotransposons and the host genome. *PLoS Genet.* **11:** e1005693.
205. Moldovan, J.B. & J.V. Moran. 2015. The zinc-finger antiviral protein ZAP inhibits LINE and *Alu* retrotransposition. *PLoS Genet.* **11:** e1005121.
206. Warren, W.C., L.W. Hillier, C. Tomlinson, *et al.* 2016. A new chicken genome assembly provides insight into avian genome structure. *G3.* doi: 10.1534/g3.116.035923.