PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Peona V *et al*. 2021 The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. *Phil. Trans. R. Soc. B* **376**: 20200186.
https://doi.org/10.1098/rstb.2020.0186

Accepted: 20 November 2020

One contribution of 12 to a theme issue 'Challenging the paradigm in sex chromosome evolution: empirical and theoretical insights with a focus on vertebrates (Part II)'.

Subject Areas:
evolution, genomics

Keywords:
transposable element, endogenous retrovirus, transcriptome, sex chromosome, W chromosome, Haldane's rule

Author for correspondence:
Valentina Peona
e-mail: valentina.peona@ebc.uu.se

†Present address: Population Ecology Group, Institute of Ecology and Evolution, Friedrich Schiller University Jena, 07743 Jena, Germany.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5450746.

THE ROYAL SOCIETY PUBLISHING

The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities

Valentina Peona[1], Octavio M. Palacios-Gimenez[1,†], Julie Blommaert[1], Jing Liu[3,4], Tri Haryoko[5], Knud A. Jønsson[6], Martin Irestedt[7], Qi Zhou[3,4,8], Patric Jern[2] and Alexander Suh[1,9]

[1]Department of Organismal Biology—Systematic Biology, and [2]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
[3]MOE Laboratory of Biosystems Homeostasis and Protection, Life Sciences Institute, Zhejiang University, Hangzhou, People's Republic of China
[4]Department of Neuroscience and Development, University of Vienna, Vienna, Austria
[5]Museum Zoologicum Bogoriense, Research Centre for Biology, Indonesian Institute of Sciences (LIPI), Cibinong, Indonesia
[6]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark
[7]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden
[8]Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310052, People's Republic of China
[9]School of Biological Sciences—Organisms and the Environment, University of East Anglia, Norwich, UK

VP, 0000-0001-5119-1837; OMP-G, 0000-0002-1472-9949; JB, 0000-0003-1411-2313; TH, 0000-0002-8549-3662; KAJ, 0000-0002-1875-9504; MI, 0000-0003-1680-6861; QZ, 0000-0002-7419-2047; PJ, 0000-0003-3393-5825; AS, 0000-0002-8979-9992

It is a broadly observed pattern that the non-recombining regions of sex-limited chromosomes (Y and W) accumulate more repeats than the rest of the genome, even in species like birds with a low genome-wide repeat content. Here, we show that in birds with highly heteromorphic sex chromosomes, the W chromosome has a transposable element (TE) density of greater than 55% compared to the genome-wide density of less than 10%, and contains over half of all full-length (thus potentially active) endogenous retroviruses (ERVs) of the entire genome. Using RNA-seq and protein mass spectrometry data, we were able to detect signatures of female-specific ERV expression. We hypothesize that the avian W chromosome acts as a refugium for active ERVs, probably leading to female-biased mutational load that may influence female physiology similar to the 'toxic-Y' effect in *Drosophila* males. Furthermore, Haldane's rule predicts that the heterogametic sex has reduced fertility in hybrids. We propose that the excess of W-linked active ERVs over the rest of the genome may be an additional explanatory variable for Haldane's rule, with consequences for genetic incompatibilities between species through TE/repressor mismatches in hybrids. Together, our results suggest that the sequence content of female-specific W chromosomes can have effects far beyond sex determination and gene dosage.

This article is part of the theme issue 'Challenging the paradigm in sex chromosome evolution: empirical and theoretical insights with a focus on vertebrates (Part II)'.

1. Introduction

Many organisms exhibit a genetic sex determination system where a pair of sex chromosomes guides sex development [1]. There are two major genetic sex-determining systems: the XY system with male heterogamety (XX females and

XY males) and the ZW system with female heterogamety (ZW females and ZZ males), whereby the Y and W are the sex-limited chromosomes (SLCs).

Sex chromosomes generally evolve from a pair of autosomes [2] that acquire a sex-determining locus and locally suppressed recombination around that locus [3,4]. The non-recombining region may remain very small, keeping the two sex chromosomes largely homomorphic. Conversely, in heteromorphic sex chromosomes, the non-recombining region may expand over time until only a small pseudo-autosomal region remains recombining, while the rest of the SLC diverges, degenerates or loses genes, and accumulates repeats [5]. The evolution of the non-recombining region of the SLC is mostly shaped by its low recombination rate. Its associated low effective population size drastically decreases the efficacy of selection [6] (i.e. accentuating the effects of drift and linked selection) and makes these chromosomes vulnerable to the accumulation of slightly deleterious mutations (e.g. through Muller's ratchet and Hill–Robertson interference mechanisms), such as repeats [3,7].

Because of their low gene content and high repeat density, SLCs were thought to not have any effect beyond sex determination and gonadal development, remaining largely understudied or even absent in the majority of the genome assemblies and studies [8]. However, recent studies on SLCs, especially in humans and other model organisms, have shown that they play roles in human diseases [9,10], male infertility [11], determining sex-specific traits [12], shaping the genome-wide heterochromatic landscape [13], exerting epistatic effects [14–16], reproductive isolation [17] and suppressing meiotic drivers on other chromosomes (e.g. through RNAi pathways) [18].

While Y chromosomes of mammals and flies have recently received considerable attention, the evolutionary implications of W chromosomes in any organism are still poorly understood. Here, we provide, to our knowledge, the first evidence that the avian W chromosome is not merely a graveyard of repetitive elements but a refugium of potentially active transposable elements (TEs) that probably have sex-specific implications. Bird genomes are known to be repeat-poor with a mean TE content of less than 10% [19], but the first female assemblies based on short [20] or long reads [21–23] showed that the non-recombining W chromosome is over 50% repetitive and especially rich in endogenous retroviruses (ERVs). By analysing reference-quality genomes of six species spanning the avian Tree of Life from both Paleognathae (emu with homomorphic sex chromosomes) and Neognathae (chicken, Anna's hummingbird, kākāpō, paradise crow, zebra finch with heteromorphic sex chromosomes), we demonstrate that the avian W has generally accumulated ERVs and probably contains active ERVs as indicated by signatures of transcription and translation of W-linked ERVs. We, therefore, hypothesize that the W is a sex-specific source of genome-wide retrotransposition and genome instability, with the male/female difference in ERVs dictating the degree of repercussions on sex differences in physiology and reproductive isolation.

2. Results and discussion

(a) Enrichment of endogenous retroviruses on the W chromosome

We analysed six avian genomes spanning the avian Tree of Life (figure 1*a*) and representing the current standard for reference-quality genome assemblies [23,26]. Autosomes had between 6 and 12% TEs on average (figure 1*b*; electronic supplementary material, table S2 and S1) and the Z chromosome had similar or slightly higher TE densities (5–17%), while the W chromosome stood out as having approximately 22–80% TEs (electronic supplementary material, table S2). Notably, we also found the homomorphic W chromosome of emu to be richer in TEs than the autosomes and Z (22 versus 6.4 and 5.6%). Generally, the Z chromosome exhibited a TE landscape more similar to the autosomes than to the W chromosome, both regarding abundances and types of TEs (figure 1*b*; electronic supplementary material, table S2). While long interspersed elements (LINEs) from the Chicken Repeat 1 (CR1) superfamily were the dominant repeats on autosomes and Z (cf. [19,27]), ERVs were the major component of the W chromosome and accounted for more than 50% of the assembled chromosome itself (electronic supplementary material, table S2).

ERVs are long terminal repeat (LTR) retrotransposons deriving from germline-inherited retrovirus integrations and exist mainly in two genomic forms [28,29]: (i) full-length elements with terminal repeats (likewise called LTRs) flanking its protein-coding genes necessary for retrotransposition; and (ii) solo-LTRs resulting from homologous recombination between the two flanking LTRs. Only full-length elements are capable of autonomous retrotransposition. Using RetroTector and LTRharvest/LTRdigest [30–32], we annotated full-length ERVs (fl-ERVs; electronic supplementary material, S2 and S3) and detected a large proportion of fl-ERVs on the W chromosome compared to the rest of the genome (figure 1*c* and table 1; electronic supplementary material, tables S3–S12). Despite the fact that the W chromosome accounted only for the 1–3% of the total length of assembled chromosomes (figure 1*c*; electronic supplementary material, table S3), this chromosome carried the same or higher numbers of fl-ERVs than the autosomes altogether, with the exception of emu with half the number on W than autosomes together (figure 1*d* and table 1). The distribution of fl-ERVs deviated significantly (χ^2-test, p-values<0.01) from a random distribution across all chromosomes (electronic supplementary material, table S4), with an impoverishment of total ERV-derived bp on the autosomes (0–0.4 times fewer bp than expected) and an extreme accumulation on the W (12–54 times more bp than expected; electronic supplementary material, table S12). By contrast, we identified a negligible amount of other full-length TEs per genome (0–11 DNA transposons, 0–8 CR1 LINEs; electronic supplementary material, table S4).

We propose a 'refugium index' (equation (4.1)) to quantify the excess accumulation of TE-derived bp on an SLC relative to the rest of the genome by comparing the observed and expected abundance of TEs. Positive values of the refugium index indicate an excess of TEs, while negative values a depletion of TEs. Because only a subset of TE copies are usually capable of (retro)transposition, we propose a 'toxicity index' as a quantitative measure for the excess of intact TE copies in the heterogametic versus homogametic sex through the presence of an SLC (equation (4.2)). The excess is calculated by comparing the number of full-length TEs in the diploid state in the two sexes. The toxicity index indicates a non-toxic SLC when equal to 0, toxicity of SLCs when positive and toxicity of Z or X when negative. The term 'toxicity' pays tribute to the recently proposed 'toxic-Y' hypothesis in *Drosophila* [13], which suggested that an excess of Y-specific active TEs can lead to male-biased

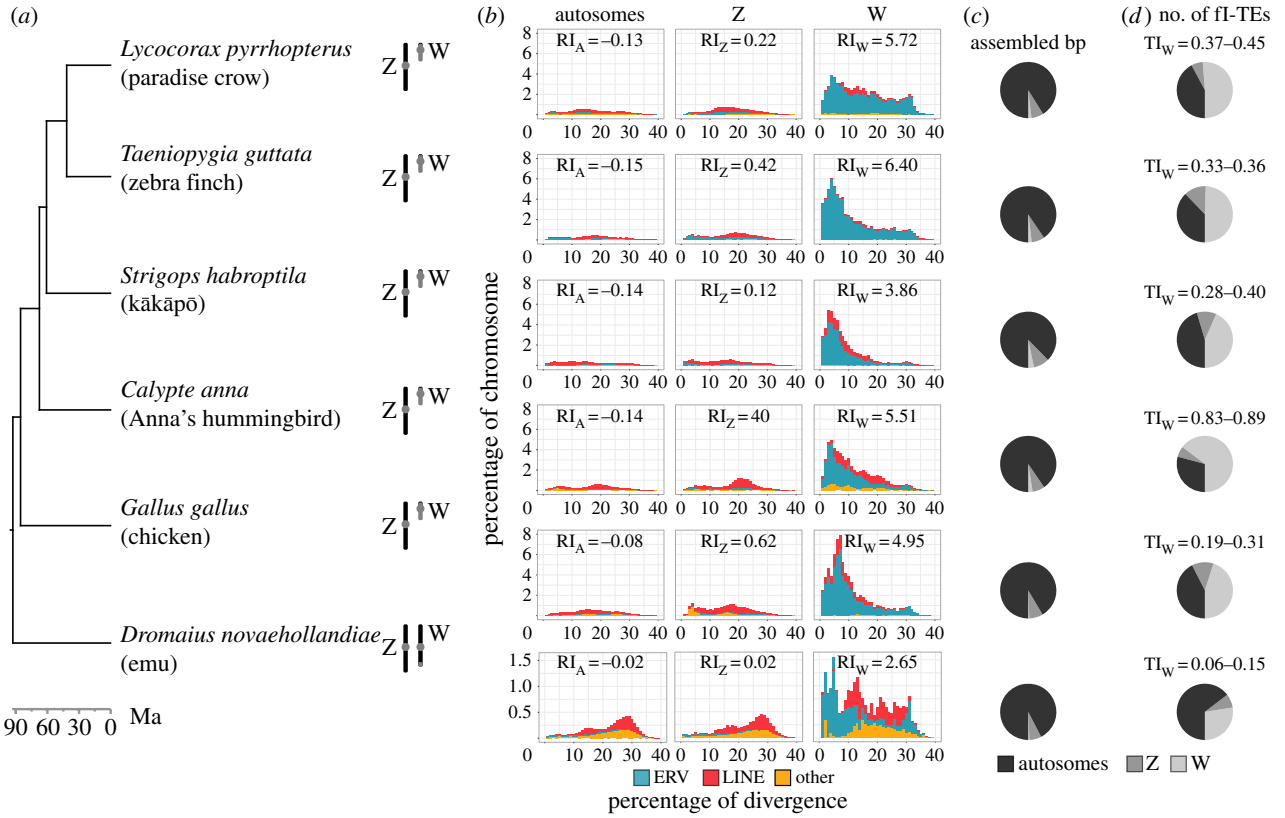

Figure 1. Massive accumulation of ERVs on W chromosomes of six female reference-quality genome assemblies spanning the avian Tree of Life. (*a*) Avian time tree after [24] with schematic homomorphic or heteromorphic sex chromosomes [25]. (*b*) TE landscapes of autosomes and sex chromosomes as stacked bar plots. Abundance of interspersed repeats (bp occupied) normalized by chromosome size plotted against percentage of divergence calculated as Kimura 2-parameter distance to consensus. The refugium index (RI) for interspersed repeats on autosomes, Z and W is indicated for each species. (*c*) Comparison of autosome and sex chromosome assembly sizes as pie charts. (*d*) Comparison of full-length TE numbers (mainly ERVs) on autosomes and sex chromosomes as pie charts. The toxicity index (TI) of the W chromosome is reported for each species as the range between estimates from RetroTector or LTRharvest + LTRdigest (table 1).

transposition and genome instability, together probably detrimental to the genome and the organism. For birds, we calculated the toxicity index as the excess of fl-ERVs carried by diploid females compared to diploid males (table 1), suggesting that females with heteromorphic sex chromosomes carried between 20 and 90% more fl-ERVs than males, and that even the emu has 7–16% more fl-ERVs in females than males despite largely homomorphic sex chromosomes [25,33]. We assume this phenomenon to reflect that the non-recombining region of the W, no matter how big or small, constantly accumulates large quantities of new TEs. It is important to note that, given the difficulties in assembling SLCs even with long-read sequencing technologies [8,23,26,34], the W chromosome models are likely to be less complete than the other chromosomes. We thus consider our W repeat annotations as well as indexes to be conservative estimates for the true repeat content.

Our results suggest that the avian W chromosome is acting as a refugium for intact and thus potentially active TEs, particularly ERVs, which may have numerous implications. We thus propose the 'refugium hypothesis' for SLCs in general: the accumulation of TEs on the SLC leads to an excess of intact TEs in the heterogametic sex, with a toxic effect absent from the SLC-lacking homogametic sex. This sex-specific toxic effect may manifest itself as sex-biased mutational load, genomic instability, ageing and genetic incompatibilities as a result of SLC-linked TE activity and heterochromatin dynamics (explained below). To quantify and test the refugium hypothesis in any sex chromosome system of interest, we introduced two indexes above: the refugium index to measure the density of TE-derived bp on the SLC relative to the remaining chromosomes; and the toxicity index to measure the number of intact TEs (i.e. full-length copies of LTRs, LINEs and DNA transposons) in the heterogametic sex relative to the other sex.

(b) Transcription and translation of W-linked endogenous retroviruses

Considering the exceptionally high number of W-linked fl-ERVs, we tested whether the avian W chromosome harbours a potentially active load of ERVs specific to females. In the absence of available retrotransposition assays for birds, we regarded the transcription and translation of W-linked ERVs as proxies of their activity. We identified W-linked single-nucleotide variants (SNVs) within ERVs by mapping genome re-sequencing data from male and female individuals, as well as female transcriptome data, to consensus sequences of our repeat library (electronic supplementary material, S5). We consider this to be a conservative subset of W-linked SNVs because we required each SNV to be present in all females and absent in all males per species. However, the paradise crow dataset that contained only one male probably gave rise to false positive W-linked SNVs. We then traced the presence of ERV proteins in the male and female proteome data available for white leghorn chicken.

We analysed zebra finch, paradise crow, chicken and emu for W-linked SNVs in genome re-sequencing and RNA-seq data mapped against ERV consensus sequences. In each species, we found between 52 and 332 ERV subfamilies with

Table 1. Number of full-length endogenous retroviruses (fl-ERVs) and other full-length TEs (LINE + DNA) found on autosomes and sex chromosomes. (fl-ERVs were identified using either RetroTector (RT) or LTRharvest + LTRdigest (LTRhd).)

species	autosomes			chromosome Z			chromosome W			males (2*n*)			females (2*n*)			W toxicity index (LTRhd)	W toxicity index (RT)
	LINE + DNA	RT	LTRhd	LINE + DNA	RT	LTRhd	LINE + DNA	RT	LTRhd	LINE + DNA	RT	LTRhd	LINE + DNA	RT	LTRhd		
Calypte anna	1	171	302	0	71	63	1	503	671	2	484	730	5	916	1338	0.834699454	0.895061728
Dromaius novaehollandiae	9	129	261	2	0	32	0	1	111	22	258	586	40	259	665	0.159539474	0.067857143
Gallus gallus	7	392	232	0	74	68	3	244	246	14	932	600	31	1102	778	0.317589577	0.197674419
Lycocorax pyrrhopterus	1	396	558	0	83	88	0	439	678	2	958	1292	4	1314	1882	0.457496136	0.372916667
Strigops habroptila	7	354	911	0	102	225	1	458	871	14	912	2272	29	1268	2918	0.289151356	0.400647948
Taeniopygia guttata	0	298	552	0	149	184	0	450	723	0	894	1472	0	1195	2011	0.366168478	0.336689038

W-linked SNVs (table 2), with ERVL subfamilies being the most represented (electronic supplementary material, table S13) and found evidence for the transcription of between 12 and 182 ERV subfamilies in female gonads or female pectoral muscle (table 2; electronic supplementary material, table S13). Our estimates of transcribed W-linked ERVs are probably just the tip of the iceberg, because we expect to identify W-linked SNVs only if those ERVs have not yet spread in the genome (e.g. very recent variants) or if they accumulated exclusively on the W chromosome (e.g. fl-ERVs only existing as solo-LTRs on other chromosomes). Alongside ERVs, we also identified W-linked SNVs in CR1 LINEs and DNA transposons (electronic supplementary material, table S13). Although there is evidence for their transcription, their scarcity of full-length elements makes CR1 LINEs and DNA transposons an unlikely source of mutational load for females.

Next, we analysed the overall RNA expression level of ERVs in male and female gonads of emu, chicken and zebra finch via RNA-seq read mapping to genomic regions annotated as LTR or ERV fragments by RepeatMasker (figure 2*a*). Overall, females expressed such ERVs more highly than males, with the single Z chromosome of females showing expression levels that matched the two male Z chromosomes. This pattern contrasts with incomplete dosage compensation of Z-linked genes in birds indicated by the usual twofold higher expression level of Z-linked genes in males [35,36]. Assuming that some of these ERVs are full-length and capable of retrotransposition, female gonads would thus be exposed to a greater mutational load. Furthermore, many of the autosomal and Z-linked ERVs showed differential expression towards females (electronic supplementary material, figure S1).

Finally, we analysed protein mass spectrometry data of white leghorn chicken gonads [37] with MaxQuant [38] for the presence of TE-related proteins and found a higher quantity of more of these proteins expressed in females than in males as indicated by a high *H*/*L* SILAC ratio (figure 2*b*; electronic supplementary material, S6). Together, these results demonstrate that some W-linked ERVs are transcribed and that females have more ERV translation than males, and that W chromosomes thus feature fl-ERVs potentially able to retrotranspose. Given our present data, we cannot distinguish whether this higher ERV translation stems solely from the W chromosome but it is plausible that the presence of an SLC causes a higher TE activity (similarly to what happens in *Drosophila* [13]).

(c) Sex-biased implications for mutational load

SLCs have been largely considered inert chromosomes with few effects beyond sex determination and gonadal development because of their low gene content (e.g. only 13 genes on *Drosophila* Y [39] and 28 genes on chicken W [21]). However, accumulating evidence shows that SLCs can have additional effects [12,40,41]. For example, it is important to highlight that the Y-linked regulatory variation within populations of *Drosophila* can have genome-wide epistatic effects [14–16,42]. This Y-linked regulatory variation cannot be explained simply by regulatory variation of the protein-coding genes and it has been proposed that the variability in Y repetitive content and structural variation are responsible for re-shaping the genome-wide heterochromatin landscape [43]. This hypothesis is known as the heterochromatin sink model, suggesting that large heterochromatin blocks on SLCs act as a sink for the heterochromatin machinery and thereby reduce the efficiency of

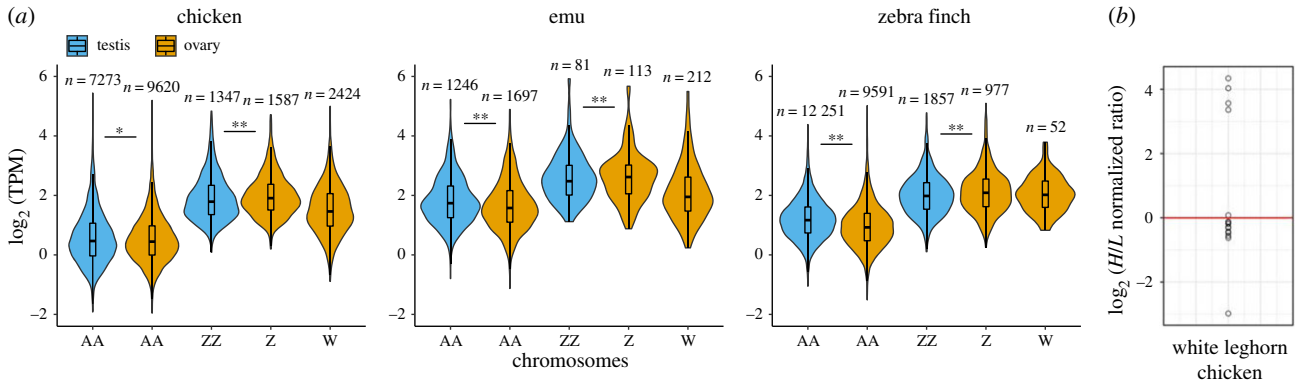

Figure 2. RNA and protein expression of ERVs in male testes and female ovaries of different birds. (*a*) RNA expression of all genomic copies/fragments (*n*) annotated as LTR or ERV by RepeatMasker. Violin plots show ERV expression levels by chromosomes using the average number of RNA-seq reads across replicates, normalized for ERVs length and library size mapping to each chromosome (A, Z and W) from ovaries (blue) and testes (orange). Significance values calculated using the Wilcoxon test (**p*-value<0.05, ***p*-value<0.01). TPM, transcripts per million reads. (*b*) Scatterplot of $\log_2$ fold-change of the *H*/*L* SILAC ratio of ERV-related peptides from 15 ERV subfamilies expressed in chicken ovaries (*H*) and testis (*L*). Values above 0 indicate proteins with female-biased expression, while values below 0 are proteins with male-biased expression. The ratios for 5 translated LINE-related peptides are found in the electronic supplementary material, S6. *H*, heavy protein labelling; *L*, light protein labelling.

Table 2. Number of female-specific and thus W-linked SNVs relative to ERV consensus sequences detected at the genomic and transcriptomic levels for each species. (More details about SNVs in ERVs, LINEs and DNA transposons are in the electronic supplementary material, table S13.)

species	no. of W-linked SNVs in ERVs	no. of ERV subfamilies with W-linked SNVs	no. of transcribed SNVs in ERVs	no. of transcribed ERV subfamilies
Dromaius novaehollandiae	764	58	82	12
Gallus gallus (red junglefowl)	2088	52	671	28
Gallus gallus (white leghorn)	6385	166	3534	102
Lycocorax pyrrhopterus	1591	198	42	21
Taeniopygia guttata	15 012	332	3306	182

heterochromatin maintenance elsewhere relative to the SLC-lacking sex [13,43].

Recently, the Y chromosome repeat content has been linked to the destabilization and loss of heterochromatin, which in turn is correlated to the shorter lifespan of the heterogametic sex [13,44]. By using *Drosophila melanogaster* experimental lines with different Y dosages (XO males, XXY females, XYY males), Brown *et al*. [13] showed that the presence and number of Y chromosomes carried are correlated with shorter lifespans. It was thus suggested that the Y itself is 'toxic' for the entire genome and organism, and this toxicity is caused by the Y-linked load of active TEs [13,45,46] whose expression is unleashed by heterochromatin loss. Possibly, the dysregulation of TEs owing to heterochromatin loss is also associated with laminopathic diseases in *Drosophila* and humans [47]. According to the refugium hypothesis proposed here, we predict that in species with a high toxicity index (i.e. excess of intact TEs on the SLCs and/or paucity thereof in the rest of the genome), this toxic effect will be more accentuated (figure 3*a*,*b*). The toxic-Y hypothesis has been recently investigated from a theoretical point of view in vertebrates with both XY and ZW systems [48] and put in contrast with the classic 'unguarded-X' hypothesis [49–51], which proposes that the expression of recessive mutations on X/Z chromosomes is the cause of the shorter lifespan in the heterogametic sex. It is important to note that reduced female lifespan in birds has been documented in many species [52–55]. Sultanova *et al*. [48] used the sizes of Y and W relative to X and Z as a proxy for toxicity, i.e. assuming that smaller SLCs are more repetitive. Although the correlation between the Y size and relative lifespan in mammals was strong, the authors did not find such a correlation for the W in birds. We note that while SLC size relative to X/Z size might indeed correlate negatively with the overall repeat content (i.e. satellites and fragmented TEs), this might not necessarily be informative for the number of intact TEs. Therefore, we propose that our toxicity index could be a more suitable proxy for toxicity because it considers the sex differences in the load of intact and (potentially) active TEs. Among the six birds compared here, emu and Anna's hummingbird would be those with the lowest and highest toxicity indexes, and it remains to be tested if this indeed is a better predictor of female lifespans.

(d) Sex-biased implications for genetic incompatibilities

In addition to TE mutational load and heterochromatin maintenance influencing organismal physiology, SLC-linked TEs can also play an important role during hybridization. This

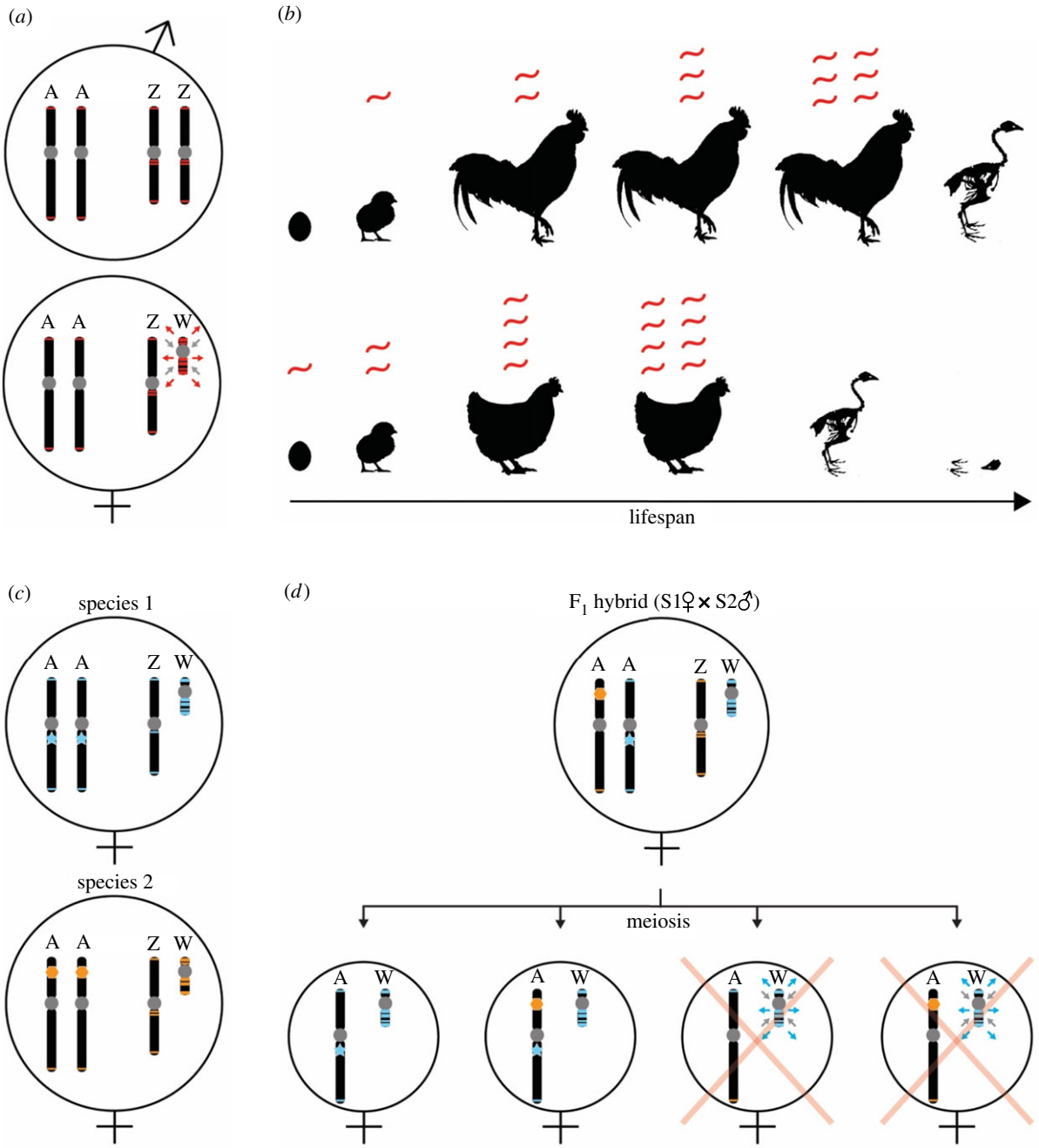

Figure 3. Synthesis of the consequences of the refugium hypothesis on micro- (*a*,*b*) and macroevolutionary (*c*,*d*) time scales. For simplicity, a schematic example of avian sex chromosomes is shown, but we expect these consequences for any ZW or XY system with SLC-linked intact TEs. (*a*) Simplified karyotypes of male and female birds indicating that the W chromosome is a TE refugium and heterochromatin sink. Grey circles: centromeres and heterochromatin blocks; red lines: intact TEs; red arrows: TEs spreading away from the W chromosome; grey arrows: heterochromatin deposition on the W chromosome. (*b*) The 'toxic' effect of the gradual de-repression of TEs during an organism's lifetime is more accentuated in females carrying more intact TEs than males. The W-linked activity of intact TEs could explain the shorter lifespan of the heterogametic sex. The toxicity of active TEs is represented by the increasing number of transcripts in red as a proxy for genome-wide TE insertions. (*c*) Simplified karyotypes of two bird species with species-specific TEs (blue and orange lines) and sequence-specific TE repressors (blue star and orange hexagon). Assuming a rapid accumulation and sequence turnover of TEs especially on the W, diverging species or populations may quickly acquire different TE/repressor repertoires. (*d*) Genetic incompatibility owing to the W chromosome in a female F_1 hybrid between species 1 and species 2 of (*c*). Schematic example of four possible meiotic products (oocytes) of the F_1 hybrid, two of which lack the blue repressor of blue TEs because of meiotic recombination between the autosomes. The TE/repressor mismatch may lead to de-repression of W-linked TEs in gametes or embryos and thereby a female-biased reduction in hybrid fitness.

point may be not overly surprising in the context of Haldane's rule, which states that upon hybridization, if there is a sterile or inviable sex, it will be the heterogametic one. Accumulating evidence suggests that hybrid genome stability can be compromised during mitosis and meiosis by species-specific differences in heterochromatin landscapes leading to uncontrolled TE activity (reviewed by Serrato-Capuchina & Matute [56]). Furthermore, species-specific families of repeats can induce lagging chromatin at cell division during early embryogenesis (when heterochromatin is first established), leading to chromosome mis-segregation and F_2 hybrid embryo death [57]. In the context of the refugium hypothesis, it is important to consider that new and active TEs are one of the main targets of heterochromatinization [58,59], and SLCs could be a source for both sex-specific and species-specific heterochromatin differences.

TEs generally evolve very rapidly in their sequence and usually only few elements remain intact and capable of transposition [60]. In addition, many TE repressor systems are in a sequence-specific arms race (e.g. piRNAs or KRAB-zinc finger proteins [58,61,62]); therefore, TE sequences and their repressors can both diverge rapidly between populations and species. Because SLCs rapidly evolve and accumulate repeats [5,18], SLCs are probably sex-specific refugia of species-specific active TEs. Hybrid incompatibility owing to TE/repressor mismatches can arise when new TE families are introduced into a naive genomic background (lacking specific repressors), which can lead to the uncontrolled proliferation of such TEs, followed by gene disruption, genome instability [63] and hybrid dysgenesis [17]. TE/repressor mismatches can already occur during meiosis in the F_1 hybrids, when recombination can separate the repressor from the controlled TEs (figure 3*c*,*d*) [64]. Although this scenario can occur in both sexes, we expect that in species with a high number of intact TEs on the SLCs relative to the rest of the genome (i.e. high toxicity index, highly heteromorphic SLCs), there are more chances for a mismatch between a repressor and intact TEs on the SLCs than for other chromosomes (figure 3*d*).

For the birds analysed here, the W chromosome is probably the main source for genome-wide new TE insertions because it contains 16–50% of all intact TEs in a diploid female. Furthermore, potential TE/repressor mismatches stemming from the W chromosome would also reinforce the observation of reduced mitochondrial (maternally inherited as the W) introgression during hybridization in birds [65]. Mitonuclear incompatibilities, i.e. mismatches between mitochondrial and nuclear alleles, play a disproportionate role in both intra-species and hybrid incompatibilities [66–69]. In ZW systems, these mitonuclear incompatibilities may be even more exacerbated because of the co-inheritance with the SLCs, and especially when SLCs feature active repeats. Thus, in addition to the preservation of dosage-sensitive genes [21,70], the W represents a reservoir of many different and intact TEs that, through their potential for de-repression in hybrids, may constitute an additional explanatory variable for Haldane's rule.

3. Conclusion

We suggest that the avian W chromosome, no matter how heteromorphic or homomorphic, is a refugium for TEs and specifically fl-ERVs, some of which are expressed and thus potentially capable of retrotransposition. This pattern should be generalizable for all birds given our broad sampling of Palaeognathae, Galloanserae and Neoaves. We propose that ERVs are continuously shaping W evolution and are one of the major contributors of structural changes of this chromosome. If so, it is reasonable to speculate that ERVs have played a relevant role in the expansion of the non-recombining region of the W (cf. [71]), for example, by contributing to the heterochromatinization of euchromatic regions through new ERV insertions.

We hope that the refugium and toxicity indexes proposed here will help testing these hypotheses in avian W chromosomes, and SLCs in general. The toxicity index measures the excess of intact TEs on an SLC, which represents the potential for genome-wide sex-specific mutational load as well as sex-specific genome instability. On the short time scale of individuals, a high toxicity index could lead to larger physiological differences between the two sexes [13]. In the long term, e.g. between populations and species, the accumulation of TEs as measured by the refugium index can have effects on reproductive isolation through TE/repressor mismatches, similar to the situation in *Drosophila* [17,57]. It is important to underline that the toxicity of SLCs should be linked to the number of intact TEs rather than to the general repetitiveness of the chromosome. Furthermore, the refugium and toxicity indexes can be useful to predict and test hybrid incompatibilities, in addition to measuring the genetic distance between nuclear and mitochondrial genes [72]. We predict that with the increasing availability of genome assemblies based on long reads, these indexes will find applicability across SLCs in general. For birds and their W chromosomes, the possible toxic effect of the W on lifespan requires additional tests *in vivo* that exclude the effects of the phenotypic sex (e.g. developing systems similar to the four core genotypes in mice [73] or the attached-X/attached-X–Y karyotypes in *Drosophila* [74,75]) and account for confounding ecological factors (e.g. intense sexual competition and predations especially of males).

To conclude, SLCs are not merely refugia for repeats with usually neutral or slightly deleterious effects on SLCs themselves, but SLC-linked intact TEs may have genome-wide effects that could effectively turn SLCs into 'toxic wastelands'.

4. Material and methods

(a) Samples, DNA, RNA and proteome data

We used the female reference-quality genome assemblies of chicken (*Gallus gallus*; GCA_000002315.5; galGal6a), paradise crow (*Lycocorax pyrrhopterus;* GCA_014706295.1) [23], emu (*Dromaius novaehollandiae;* GCA_016128335.1) [76], Anna's hummingbird (*Calypte anna*; GCA_003957555.2; bCalAnn1_v1.p) [26], kākāpō (*Strigops habroptila*; GCA_004027225.2; bStrHab1.2.pri) [26] and zebra finch (*Taeniopygia guttata;* GCA_009859065.2; bTaeGut2.pri.v2) [26]. All these six assemblies have chromosome models and we carried out all analyses considering only using assembled chromosomes, i.e. discarding unplaced contigs and scaffolds.

For chicken, Illumina genome re-sequencing libraries were collected for two females and three males of *Gallus gallus gallus* (red junglefowl) from [77] (originally uploaded on NCBI as of undetermined sex) and a female library of *Gallus gallus bankiva* (red junglefowl from Java) from [78]. The sexes of the individuals from [77] were determined using the SEXCMD with default sex markers [79]. Red junglefowl RNA-seq libraries of a female (ovary) and of a male (testes) were retrieved from [80]. We also collected publicly available data for the chicken breed white leghorn, i.e. Illumina genome re-sequencing libraries of one female and three males from [78,81], RNA-seq libraries and protein mass spectrometry libraries for five ovaries and five testes [37].

For paradise crow, we used one 10X Genomics Chromium linked-read library of DNA from a pectoral muscle sample of a female from [23]. We also newly generated such data for three females and one male using the same methods [23] and generated RNA-seq data from female pectoral muscle (preserved in RNAlater). RNA was extracted with phenol-based phase separation using the TRIzol reagent (ThermoFisher Scientific) following the standard protocol recommended by the supplier, followed by DNase treatment for 30 min using the DNA-free DNA removal kit (ThermoFisher Scientific). Sequencing libraries were prepared according to the TruSeq stranded total library preparation kit with RiboZero Gold treatment (Illumina, Inc., cat no. 20020598/9). Paired-reads (150 bp) were sequenced on the NovaSeq SP flowcell (Illumina, Inc.).

For zebra finch, we used Illumina genome re-sequencing libraries of four females and four males from [82], and RNA-seq libraries of two ovaries and one testis from [83,84].

Finally, for emu, we collected Illumina genome re-sequencing libraries of two females and two males from [85–87], and RNA-seq libraries for seven ovaries and five testes from [86,88].

More details and accession numbers for all the libraries and genomic sequences used here can be found in the electronic supplementary material, table S1.

(b) Repeat annotation

To best annotate repeats in all six avian species, we made sure to have species-specific repeat predictions for each. The repeat libraries of chicken, paradise crow and zebra finch were already manually curated elsewhere [23,89,90] while species-specific repeat libraries did not exist for emu, Anna's hummingbird and kākāpō. Therefore, we de novo characterized repetitive elements in these last three species using RepeatModeler2 [91] and manually curated those sequences labelled as 'LTR' and 'unknown' following the same method as in [23]. We also inspected consensus sequences with unusual classification for being avian repeats like many DNA transposon superfamilies [19]. We then concatenated the newly curated libraries with the avian consensus sequences from Repbase [92], hooded crow [93], blue-capped cordon bleu [94], collared flycatcher [95] and paradise crow [23], and used this final library to mask all six genomes with RepeatMasker [96]. The new repeat libraries and notes on their classification are given in the electronic supplementary material, S7.

(c) Quantity of endogenous retrovirus transcription and their differential expression

We used Illumina RNA-seq reads from adult gonads from emu, chicken and zebra finch (electronic supplementary material, table S1) mapped against genomic copies/fragments annotated as LTR or ERV by RepeatMasker to quantify ERV transcription levels and investigate whether the ERVs were differentially expressed across available tissues. For these species, three to five biological replicates for every tissue were used.

Raw RNA-seq data were quality controlled using FastQC [97] and trimmed with TrimGalore [98] using default settings, then mapped to the respective reference genomes using STAR [99]. The alignment was filtered by running featureCounts function from the package Subread v2.0.0 in paired-end mode [100], and only uniquely mapping reads were retained. We provided featureCounts with a filtered RepeatMasker .out file containing only repeat copies annotated as LTR or ERV. Per-genome counts were obtained using read counts and lengths of corresponding ERVs. DESeq2 1.20.0 [101] implemented in the R Bioconductor package was used for relative quantification of the ERV transcripts and for calculating the TPM (transcripts per million), giving a normalized ERV expression level. Male reads that mapped to the W chromosome represented low counts and were, therefore, removed during the normalization step. The values from replicates of each sample were averaged for the final plots of ERV expression. To identify biased ERVs per chromosome type (i.e. autosomes, Z and W), we compared adult gonads from male and female individuals. The statistical analysis of differentially expressed ERVs was performed using DESeq2. All p-values were adjusted ($p$adj) using the Wald test. The degree of bias was determined by the $\log_2$ fold-change ($\log_2$FC) difference between conditions. Therefore, the ERVs with $\log_2\text{FC} > 0$ and $\log_2\text{FC} < 0$ together with a $p\text{adj} < 0.05$ were considered as biased ERVs in the conditions.

(d) Full-length transposable element detection and abundance

Here, we define full-length TEs as possible (retro)transposition-competent elements with relatively complete structures and the potential to produce transcripts. We identified fl-TEs in all the six avian genomes by adopting different methods for DNA transposons, LINEs (e.g. CR1) and LTR retrotransposons (ERVs). For DNA transposons and LINEs, we first identified open reading frames (ORFs) in the insertions annotated by RepeatMasker, then translated such ORFs and aligned with RPS-BLAST [102] against a custom Pfam [103] database containing transposon-related proteins (similar approach to [104]). ORFs from LINEs of at least 600 bp that spanned 90% of both endonuclease and reverse transcriptase domains were considered as full-length elements. Likewise, ORFs belonging to DNA transposons of at least 1 kb that spanned 90% of the transposase protein domain were considered full length.

In order to detect and quantify fl-ERVs, we used RetroTector [30] as well as LTRharvest [31] together with LTRdigest [32]. RetroTector results were filtered for scores over 300 and presence of 5′-LTR and 3′-LTR, as well as ORFs with complete or partly complete *gag*, *pol* and *env* genes as previously described in [30,105]. LTRharvest results were filtered for false positive using LTRdigest in combination with hidden Markov models profiles of LTR retrotransposon-related proteins downloaded from Pfam [103] and GyDB [106].

(e) Identification of single-nucleotide variants of W-linked endogenous retroviruses and their transcription and translation

To verify the hypothesis that the W chromosome is a refugium of intact and potentially active ERVs, we identified W-linked SNVs within ERVs and traced their transcription in RNA-seq data and translation in protein mass spectrometry data wherever possible. W-linked ERV transcription was analysed in *G. gallus*, *L. pyrrhopterus* and *T. guttata* (electronic supplementary material, table S1). ERV translation was analysed in *G. gallus* white leghorn breed [37]. RNA-seq and proteome libraries selected for this analysis were from gonad tissue with the exception of *L. pyrrhopterus* for which the RNA-seq data were generated from female pectoral muscle.

To identify W-linked SNVs from male/female read mapping, we used the WhatGene pipeline developed by Ruiz-Ruano *et al*. [107] for SNV analyses of B chromosomes and germline-restricted chromosomes [108] where we mapped male and female genome re-sequencing reads to the consensus sequences of our repeat library. We considered variants to be W-linked if they were present in all females but absent in males. We then checked for the presence of these W-linked variants in the RNA-seq data always following the WhatGene pipeline. Variants that were called W-linked from genomic data but were present in male transcriptomic data were discarded as false positives owing to sample size.

To check for the presence of ERV-related proteins in white leghorn chicken proteome data, we extracted the ORFs from ERV consensus sequences and translated them into peptides using ORFFinder [109]. The peptide sequences were used as query database for MaxQuant 1.6.17.0 [38]. We used the experimental parameters described in [37] (electronic supplementary material, S5); search results were filtered with a false discovery rate of 0.01. Second peptides, dependent peptides and match between runs parameters were enabled.

(f) Refugium index and toxicity index

To test whether intact TEs are uniformly distributed throughout the genome, we compared the observed total number of fl-ERVs (assuming that the numbers of other intact TEs are negligible in avian genomes [19]) on autosomes and sex chromosome to their expected values with a χ^2-test with 2 degrees of freedom. We calculated the expected values of TE densities on the chromosomes by assuming a uniform density of these elements across chromosomes (electronic supplementary material, tables S3 and S4). Next, we calculated the refugium and toxicity indexes, which are described below for SLCs in general.

The refugium index (equation (4.1)) calculates the percentage of excess or depletion of observed TE-derived bp ($\%TE_{obs}$) with respect to the genome-wide average of the total TE-derived bp of a haploid genome assembly ($\%TE_{exp}$). We recommend estimating TE densities in RepeatMasker or similar homology-based annotations using a species-specific repeat library combined with libraries of related species in Repbase or similar databases:

$$\text{refugium index} = \frac{\%TE_{obs} - \%TE_{exp}}{\%TE_{exp}}. \quad (4.1)$$

The refugium index indicates whether an SLC shows an excess (RI > 0) or a depletion of TEs (RI < 0). Furthermore, the refugium index can be estimated for any chromosome of interest, considering all TEs together or specific TE groups separately.

The toxicity index (equation (4.2)) calculates the excess of intact TEs present in the heterogametic sex with respect to the homogametic sex. Here, $2n_{hom}$ and $2n_{het}$ are the total numbers of intact TEs in the diploid state in the homogametic sex (2 × autosomes + 2 × Z or X) and the heterogametic sex (2 × autosomes + 1 × Z or X + 1 × W or Y), respectively. We recommend quantifying intact TEs as the sum of the number of full-length LTR retrotransposons (incl. ERVs) in RetroTector/LTRharvest or similar structure-based approaches and the number of copies spanning greater than 90% of the ORFs of DNA transposons (i.e. transposase) and LINEs (i.e. ORF1 or ORF2) in RPS-BLAST or similar homology-based searches:

$$\text{toxicity index} = \frac{2n_{het} - 2n_{hom}}{2n_{hom}}. \quad (4.2)$$

The toxicity index indicates whether there is no sex difference in toxicity (TI = 0), toxicity of the W or Y chromosome (TI > 0) or even toxicity of the Z or X chromosome (TI < 0). Consequently, we expect the toxicity index to be applicable not only to XY and ZW systems, but also XO systems.

Ethics. All research in Indonesia was carried out in compliance with the ethics guidelines supplied by the Research Center for Biology, Indonesian Institute of Sciences (RCB-LIPI), the Bogor Zoological Museum and the State Ministry of Research and Technology (RISTEK) under research permit numbers: 491/SIP/FRP/SM/XI/2013, 420/SIP/FRP/SM/XI/2013 and 421/SIP/FRP/SM/X/2013.

Data accessibility. All the data are publicly available, and the code is accessible at http://github.com/ValentinaBOP/Wrefugium. All newly generated data were deposited in BioProject PRJNA604967.

Authors' contributions. V.P. and A.S. designed the study, wrote and revised the subsequent drafts with input from all authors. V.P. analysed the data and wrote the first manuscript draft. K.A.J. and T.H. provided paradise crow samples. M.I. extracted paradise crow DNA. O.M.P.-G. extracted paradise crow RNA and helped with the analysis of repeat transcription and their differential expression. J.B. helped with repeat annotation. P.J. ran RetroTector analyses. Q.Z. and J.L. provided the emu genome assembly. A.S. supervised the study. All authors read and approved the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This research was supported by grants from the Swedish Research Council Formas (2017-01597 to A.S.; 2018-01008 to P.J.), the Swedish Research Council Vetenskapsrådet (2016-05139 to A.S.; 2018-03017 to P.J.; 621-2014-5113 and 2019-03900 to M.I.), the SciLife-Lab Swedish Biodiversity Program (2015-R14 to A.S.), Villum Foundation (Young Investigator Programme, project no. 15560 to K.A.J.) and from the Carlsberg Foundation (Distinguished Associate Professor Fellowship, project no. CF17-0248 to K.A.J.). K.A.J. acknowledges a National Geographic Research and Exploration grant (8853-10), the Dybron Hoffs Foundation and the Corrit Foundation for financial support for fieldwork in Indonesia.

Acknowledgements. We thank Francisco Ruiz-Ruano for help with running WhatGene, Philipp Pottmeier for help with RNA extractions, Alexander J. Charles for help with running MaxQuant, and the Suh laboratory and Johannesson laboratory for helpful discussions. We thank Max Käller, Phil Ewels, Remi-André Olsen, Joel Gruselius and Fanny Taborsak-Lines for generating 10X data and assemblies at SciLifeLab Stockholm. We thank Erich Jarvis and the Vertebrate Genomes Project (VGP) for making their zebra finch, kākāpō and hummingbird assemblies available prior to publication, and Guojie Zhang and the 10 000 Bird Genomes (B10K) project for doing the same with emu short read data. We thank Marco Ricci, Ivar Westerberg, Jesper Boman, Diem Nguyen and three anonymous reviewers for their comments on the manuscript. The Swedish Biodiversity Program has been made available by support from the Knut and Alice Wallenberg Foundation. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, which is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory, and by the National Genomics Infrastructure in Stockholm. Both facilities are funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). We thank the State Ministry of Research and Technology (RISTEK); the Ministry of Forestry, Republic of Indonesia; the Research Center for Biology, Indonesian Institute of Sciences (RCB-LIPI); the Bogor Zoological Museum for providing permits to carry out fieldwork in Indonesia and to export select samples; the Natural Resources and Conservation Agency (BKSDA) Maluku, Ministry of Environment and Forestry-Republic of Indonesia.

References

1. Bachtrog D *et al.* 2014 Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899. (doi:10.1371/journal.pbio.1001899)
2. Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE. 2020 Sex chromosome evolution: so many exceptions to the rules. *Genome Biol. Evol.* **12**, 750–763. (doi:10.1093/gbe/evaa081)
3. Charlesworth B, Charlesworth D. 1978 A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997. (doi:10.1086/283342)
4. Rice WR. 1987 The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution (NY)* **41**, 911–914. (doi:10.2307/2408899)
5. Mank JE. 2012 Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. *Chromosome Res.* **20**, 21–33. (doi:10.1007/s10577-011-9251-2)
6. Bachtrog D. 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124. (doi:10.1038/nrg3366)
7. Wright AE, Dean R, Zimmer F, Mank JE. 2016 How to make a sex chromosome. *Nat. Commun.* **7**, 12087. (doi:10.1038/ncomms12087)
8. Tomaszkiewicz M, Medvedev P, Makova KD. 2017 Y and W chromosome assemblies: approaches and discoveries. *Trends Genet.* **33**, 266–282. (doi:10.1016/j.tig.2017.01.008)
9. Charchar FJ *et al.* 2012 Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *Lancet* **379**, 915–922. (doi:10.1016/S0140-6736(11)61453-0)
10. Kido T, Lau Y-FC. 2015 Roles of the Y chromosome genes in human cancers. *Asian J. Androl.* **17**, 373–380. (doi:10.4103/1008-682X.150842)
11. Dhanoa JK, Mukhopadhyay CS, Arora JS. 2016 Y-chromosomal genes affecting male fertility: a review. *Vet. World* **9**, 783–791. (doi:10.14202/vetworld.2016.783-791)
12. Stoltenberg SF, Hirsch J. 1997 Y-chromosome effects on *Drosophila* geotaxis interact with genetic or cytoplasmic background. *Anim. Behav.* **53**, 853–864. (doi:10.1006/anbe.1996.0351)
13. Brown EJ, Nguyen AH, Bachtrog D. 2020 The Y chromosome may contribute to sex-specific ageing

in *Drosophila*. *Nat. Ecol. Evol.* **4**, 853–862. (doi:10.1038/s41559-020-1179-5)

14. Jiang P-P, Hartl DL, Lemos B. 2010 Y not a dead end: epistatic interactions between Y-linked regulatory polymorphisms and genetic background affect global gene expression in *Drosophila melanogaster*. *Genetics* **186**, 109–118. (doi:10.1534/genetics.110.118109)
15. Lemos B, Branco AT, Hartl DL. 2010 Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proc. Natl Acad. Sci. USA* **107**, 15 826–15 831. (doi:10.1073/pnas.1010383107)
16. Kutch IC, Fedorka KM. 2017 A test for Y-linked additive and epistatic effects on surviving bacterial infections in *Drosophila melanogaster*. *J. Evol. Biol.* **30**, 1400–1408. (doi:10.1111/jeb.13118)
17. Kidwell MG, Kidwell JF, Sved JA. 1977 Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* **86**, 813–833. (doi:10.1093/genetics/86.4.813)
18. Bachtrog D. 2020 The Y chromosome as a battleground for intragenomic conflict. *Trends Genet.* **36**, 510–522. (doi:10.1016/j.tig.2020.04.008)
19. Kapusta A, Suh A. 2017 Evolution of bird genomes—a transposon's-eye view. *Ann. N.Y. Acad. Sci.* **1389**, 164–185. (doi:10.1111/nyas.13295)
20. Smeds L *et al.* 2015 Evolutionary analysis of the female-specific avian W chromosome. *Nat. Commun.* **6**, 7330. (doi:10.1038/ncomms8330)
21. Bellott DW *et al.* 2017 Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat. Genet.* **49**, 387–394. (doi:10.1038/ng.3778)
22. Warren WC *et al.* 2017 A new chicken genome assembly provides insight into avian genome structure. *G3 Genes|Genomes|Genetics* **7**, 109–117. (doi:10.1534/g3.116.035923)
23. Peona V *et al.* 2020 Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol. Ecol. Resour.* **21**, 263–286. (doi:10.1111/1755-0998.13252)
24. Claramunt S, Cracraft J. 2015 A new time tree reveals Earth history's imprint on the evolution of modern birds. *Sci. Adv.* **1**, e1501005. (doi:10.1126/sciadv.1501005)
25. Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MTP, Zhang G. 2014 Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science (80–.)* **346**, 1246338. (doi:10.1126/science.1246338)
26. Rhie A *et al.* 2021 Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746. (doi:10.1038/s41586-021-03451-0)
27. Zhang G *et al.* 2014 Comparative genomics reveals insights into avian genome evolution and adaptation. *Science (80–.)* **346**, 1311–1320. (doi:10.1126/science.1251385)
28. Kent TV, Uzunović J, Wright SI. 2017 Coevolution between transposable elements and recombination. *Phil. Trans. R. Soc. B* **372**, 20160458. (doi:10.1098/rstb.2016.0458)
29. Jedlicka P, Lexa M, Kejnovsky E. 2020 What can long terminal repeats tell us about the age of LTR retrotransposons, gene conversion and ectopic recombination? *Front. Plant Sci.* **11**, 644. (doi:10.3389/fpls.2020.00644)
30. Sperber GO, Airola T, Jern P, Blomberg J. 2007 Automated recognition of retroviral sequences in genomic data—RetroTector©. *Nucleic Acids Res.* **35**, 4964–4976. (doi:10.1093/nar/gkm515)
31. Ellinghaus D, Kurtz S, Willhoeft U. 2008 LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinf.* **9**, 18. (doi:10.1186/1471-2105-9-18)
32. Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009 Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res.* **37**, 7002–7013. (doi:10.1093/nar/gkp759)
33. Ogawa A, Murata K, Mizuno S. 1998 The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *Proc. Natl Acad. Sci. USA* **95**, 4415–4418. (doi:10.1073/pnas.95.8.4415)
34. Peona V, Weissensteiner MH, Suh A. 2018 How complete are 'complete' genome assemblies?—an avian perspective. *Mol. Ecol. Resour.* **18**, 1188–1195. (doi:10.1111/1755-0998.12933)
35. Itoh Y, Replogle K, Kim Y-H, Wade J, Clayton DF, Arnold AP. 2010 Sex bias and dosage compensation in the zebra finch versus chicken genomes: general and specialized patterns among birds. *Genome Res.* **20**, 512–518. (doi:10.1101/gr.102343.109)
36. Adolfsson S, Ellegren H. 2013 Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. *Mol. Biol. Evol.* **30**, 806–810. (doi:10.1093/molbev/mst009)
37. Uebbing S, Konzer A, Xu L, Backström N, Brunström B, Bergquist J, Ellegren H. 2015 Quantitative mass spectrometry reveals partial translational regulation for dosage compensation in chicken. *Mol. Biol. Evol.* **32**, 2716–2725. (doi:10.1093/molbev/msv147)
38. Tyanova S, Temu T, Cox J. 2016 The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319. (doi:10.1038/nprot.2016.136)
39. Griffin RM, Le Gall D, Schielzeth H, Friberg U. 2015 Within-population Y-linked genetic variation for lifespan in *Drosophila melanogaster*. *J. Evol. Biol.* **28**, 1940–1947. (doi:10.1111/jeb.12708)
40. Chippindale AK, Rice WR. 2001 Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **98**, 5677–5682. (doi:10.1073/pnas.101456898)
41. Werner RJ, Schultz BM, Huhn JM, Jelinek J, Madzo J, Engel N. 2017 Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biol. Sex Differ.* **8**, 28. (doi:10.1186/s13293-017-0150-x)
42. Case LK *et al.* 2013 The Y chromosome as a regulatory element shaping immune cell transcriptomes and susceptibility to autoimmune disease. *Genome Res.* **23**, 1474–1485. (doi:10.1101/gr.156703.113)
43. Francisco FO, Lemos B. 2014 How do Y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. *J. Genomics* **2**, 94–103. (doi:10.7150/jgen.8043)
44. Maklakov AA, Lummaa V. 2013 Evolution of sex differences in lifespan and aging: causes and constraints. *Bioessays* **35**, 717–724. (doi:10.1002/bies.201300021)
45. Wei K, Gibilisco L, Bachtrog D. 2020 Epigenetic conflict on a degenerating Y chromosome increases mutational burden in *Drosophila* males. *Nat. Commun.* **11**, 5537. (doi:10.1101/2020.07.19.210948)
46. Nguyen AH, Bachtrog D. 2020 Toxic Y chromosome: increased repeat expression and age-associated heterochromatin loss in male *Drosophila* with a young Y chromosome. *bioRxiv* 2020.07.21.214528. (doi:10.1101/2020.07.21.214528)
47. Andrenacci D, Cavaliere V, Lattanzi G. 2020 The role of transposable elements activity in aging and their possible involvement in laminopathic diseases. *Ageing Res. Rev.* **57**, 100995. (doi:10.1016/j.arr.2019.100995)
48. Sultanova Z, Downing PA, Carazo P. 2020 Genetic sex determination and sex-specific lifespan in tetrapods—evidence of a toxic Y effect. *bioRxiv* 2020.03.09.983700. (doi:10.1101/2020.03.09.983700)
49. Trivers R. 1985 *Social evolution*. Menlo Park, CA: Benjamin Cummings Publishing Co.
50. Xirocostas ZA, Everingham SE, Moles AT. 2020 The sex with the reduced sex chromosome dies earlier: a comparison across the tree of life. *Biol. Lett.* **16**, 20190867. (doi:10.1098/rsbl.2019.0867)
51. Brengdahl M, Kimber CM, Maguire-Baxter J, Friberg U. 2018 Sex differences in life span: females homozygous for the X chromosome do not suffer the shorter life span predicted by the unguarded X hypothesis. *Evolution* (*NY*) **72**, 568–577. (doi:10.1111/evo.13434)
52. Lambertucci SA, Carrete M, Donázar JA, Hiraldo F. 2012 Large-scale age-dependent skewed sex ratio in a sexually dimorphic avian scavenger. *PLoS ONE* **7**, e46347. (doi:10.1371/journal.pone.0046347)
53. Clutton-Brock TH, Isvaran K. 2007 Sex differences in ageing in natural populations of vertebrates. *Proc. R. Soc. B* **274**, 3097–3104. (doi:10.1098/rspb.2007.1138)
54. Pipoly I, Bókony V, Kirkpatrick M, Donald PF, Székely T, Liker A. 2015 The genetic sex-determination system predicts adult sex ratios in tetrapods. *Nature* **527**, 91–94. (doi:10.1038/nature15380)
55. Donald PF. 2007 Adult sex ratios in wild bird populations. *Ibis* (*Lond. 1859*) **149**, 671–692. (doi:10.1111/j.1474-919X.2007.00724.x)
56. Serrato-Capuchina A, Matute DR. 2018 The role of transposable elements in speciation. *Genes* (*Basel*) **9**, 254. (doi:10.3390/genes9050254)
57. Ferree PM, Barbash DA. 2009 Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*.

PLoS Biol. **7**, e1000234. (doi:10.1371/journal.pbio.1000234)

58. Kelleher ES, Barbash DA, Blumenstiel JP. 2020 Taming the turmoil within: new insights on the containment of transposable elements. *Trends Genet.* **36**, 474–489. (doi:10.1016/j.tig.2020.04.007)
59. Choi JY, Lee YCG. 2020 Double-edged sword: the evolutionary consequences of the epigenetic silencing of transposable elements. *PLoS Genet.* **16**, e1008872. (doi:10.1371/journal.pgen.1008872)
60. Huang CRL, Burns KH, Boeke JD. 2012 Active transposition in genomes. *Annu. Rev. Genet.* **46**, 651–675. (doi:10.1146/annurev-genet-110711-155616)
61. Yang P, Wang Y, Macfarlan TS. 2017 The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet.* **33**, 871–881. (doi:10.1016/j.tig.2017.08.006)
62. Czech B, Munafò M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, Hannon GJ. 2018 piRNA-guided genome defense: from biogenesis to silencing. *Annu. Rev. Genet.* **52**, 131–157. (doi:10.1146/annurev-genet-120417-031441)
63. Dion-Côté A-M, Barbash DA. 2017 Beyond speciation genes: an overview of genome stability in evolution and speciation. *Curr. Opin. Genet. Dev.* **47**, 17–23. (doi:10.1016/j.gde.2017.07.014)
64. Rogers RL. 2015 Chromosomal rearrangements as barriers to genetic homogenization between archaic and modern humans. *Mol. Biol. Evol.* **32**, 3064–3078. (doi:10.1093/molbev/msv204)
65. Petit RJ, Excoffier L. 2009 Gene flow and species delimitation. *Trends Ecol. Evol.* **24**, 386–393. (doi:10.1016/j.tree.2009.02.011)
66. Fishman L, Willis JH. 2006 A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution* (*NY*) **60**, 1372–1381. (doi:10.1111/j.0014-3820.2006.tb01216.x)
67. Trier CN, Hermansen JS, Sætre G-P, Bailey RI. 2014 Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. *PLoS Genet.* **10**, e1004075. (doi:10.1371/journal.pgen.1004075)
68. Haddad R, Meter B, Ross JA. 2018 The genetic architecture of intra-species hybrid mito-nuclear epistasis. *Front. Genet.* **9**, 481. (doi:10.3389/fgene.2018.00481)
69. Lima TG, Burton RS, Willett CS. 2019 Genomic scans reveal multiple mito-nuclear incompatibilities in population crosses of the copepod *Tigriopus californicus*. *Evolution* (*NY*) **73**, 609–620. (doi:10.1111/evo.13690)
70. Bellott DW, Page DC. 2021 Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Res.* **31**, 198–210. (doi:10.1101/gr.268516.120)
71. Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018 Why do sex chromosomes stop recombining? *Trends Genet.* **34**, 492–503. (doi:10.1016/j.tig.2018.04.001)
72. Allen R *et al.* 2020 A mitochondrial genetic divergence proxy predicts the reproductive compatibility of mammalian hybrids. *Proc. R. Soc. B* **287**, 20200690. (doi:10.1098/rspb.2020.0690)
73. Arnold AP. 2009 Mouse models for evaluating sex chromosome effects that cause sex differences in non-gonadal tissues. *J. Neuroendocrinol.* **21**, 377–386. (doi:10.1111/j.1365-2826.2009.01831.x)
74. Morgan LV. 1925 Polyploidy in *Drosophila melanogaster* with two attached X chromosomes. *Genetics* **10**, 148–178. (doi:10.1093/genetics/10.2.148)
75. Green MM, Piergentili R. 2000 On the origin of metacentric, attached-X (A-X) chromosomes in *Drosophila melanogaster* males. *Proc. Natl Acad. Sci. USA* **97**, 14 484–14 487. (doi:10.1073/pnas.250483497)
76. Liu J *et al.* 2021 A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. *Genome Res.* **31**, 497–511. (doi:10.1101/gr.271569.120)
77. Wang M-S *et al.* 2015 Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol. Biol. Evol.* **32**, 1880–1889. (doi:10.1093/molbev/msv071)
78. Piégu B *et al.* 2020 Variations in genome size between wild and domesticated lineages of fowls belonging to the *Gallus gallus* species. *Genomics* **112**, 1660–1673. (doi:10.1016/j.ygeno.2019.10.004)
79. Jeong S, Kim J, Park W, Jeon H, Kim N. 2017 SEXCMD: development and validation of sex marker sequences for whole-exome/genome and RNA sequencing. *PLoS ONE* **12**, e0184087. (doi:10.1371/journal.pone.0184087)
80. McCarthy FM *et al.* 2019 Chickspress: a resource for chicken gene expression. *Database* **2019**, baz058. (doi:10.1093/database/baz058)
81. Oh D, Son B, Mun S, Oh MH, Oh S, Ha J, Yi J, Lee S, Han K. 2016 Whole genome re-sequencing of three domesticated chicken breeds. *Zool. Sci.* **33**, 73–77. (doi:10.2108/zs150071)
82. Singhal S *et al.* 2015 Stable recombination hotspots in birds. *Science (80–.)* **350**, 928–932. (doi:10.1126/science.aad0843)
83. Biederman MK, Nelson MM, Asalone KC, Pedersen AL, Saldanha CJ, Bracht JR. 2018 Discovery of the first germline-restricted gene by subtractive transcriptomic analysis in the zebra finch, *Taeniopygia guttata*. *Curr. Biol.* **28**, 1620–1627.e5. (doi:10.1016/j.cub.2018.03.067)
84. Yin Z-T *et al.* 2019 Revisiting avian 'missing' genes from de novo assembled transcripts. *BMC Genomics* **20**, 4. (doi:10.1186/s12864-018-5407-1)
85. Vicoso B, Kaiser VB, Bachtrog D. 2013 Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl Acad. Sci. USA* **110**, 6453–6458. (doi:10.1073/pnas.1217027110)
86. Sackton TB *et al.* 2019 Convergent regulatory evolution and loss of flight in paleognathous birds. *Science (80–.)* **364**, 74–78. (doi:10.1126/science.aat7244)
87. Feng S *et al.* 2020 Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257.
88. Xu L, Wa Sin SY, Grayson P, Edwards SV, Sackton TB. 2019 Evolutionary dynamics of sex chromosomes of paleognathous birds. *Genome Biol. Evol.* **11**, 2376–2390. (doi:10.1093/gbe/evz154)
89. Wicker T *et al.* 2005 The repetitive landscape of the chicken genome. *Genome Res.* **15**, 126–136. (doi:10.1101/gr.2438004)
90. Warren WC *et al.* 2010 The genome of a songbird. *Nature* **464**, 757–762. (doi:10.1038/nature08819)
91. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020 RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457. (doi:10.1073/pnas.1921046117)
92. Bao W, Kojima KK, Kohany O. 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11. (doi:10.1186/s13100-015-0041-9)
93. Weissensteiner MH *et al.* 2020 Discovery and population genomics of structural variation in a songbird genus. *Nat. Commun.* **11**, 3403. (doi:10.1038/s41467-020-17195-4)
94. Boman J, Frankl-Vilches C, da Silva dos Santos M, de Oliveira EHC, Gahr M, Suh A. 2019 The genome of blue-capped cordon-bleu uncovers hidden diversity of LTR retrotransposons in zebra finch. *Genes* **10**, 301. (doi:10.3390/genes10040301)
95. Suh A, Smeds L, Ellegren H. 2018 Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. *Mol. Ecol.* **27**, 99–111. (doi:10.1111/mec.14439)
96. Smit AFA, Hubley R, Green P. 2015 RepeatMasker Open-4.0. See http://www.repeatmasker.org.
97. Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S. 2015 FastQC: a quality control tool for high throughput sequence data. *Babraham Inst.* **1**, 1. See https://www.bioinformatics.babraham.ac.uk/projects/fastqc.
98. Krueger F. 2015 Trim galore. A wrapper tool around Cutadapt FastQC to consistently apply Qual. Adapt. trimming to FastQ files 516, 517. See https://www.bioinformatics.babraham.ac.uk/projects/trim_galore.
99. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (doi:10.1093/bioinformatics/bts635)
100. Liao Y, Smyth GK, Shi W. 2019 The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47. (doi:10.1093/nar/gkz114)
101. Love MI, Huber W, Anders S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. (doi:10.1186/s13059-014-0550-8)
102. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST+: architecture and applications. *BMC Bioinf.* **10**, 421. (doi:10.1186/1471-2105-10-421)
103. El-Gebali S *et al.* 2018 The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432. (doi:10.1093/nar/gky995)

104. Galbraith JD, Ludington AJ, Suh A, Sanders KL, Adelson DL. 2020 New environment, new invaders—repeated horizontal transfer of LINEs to sea snakes. *Genome Biol. Evol.* **12**, 2370–2383. (doi:10.1093/gbe/evaa208)
105. Hayward A, Cornwallis CK, Jern P. 2015 Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc. Natl Acad. Sci. USA* **112**, 464–469. (doi:10.1073/pnas.1414980112)
106. Llorens C *et al.* 2010 The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* **39**, D70–D74. (doi:10.1093/nar/gkq1061)
107. Ruiz-Ruano FJ, Navarro-Domínguez B, López-León MD, Cabrero J, Camacho JPM. 2019 Evolutionary success of a parasitic B chromosome rests on gene content. *bioRxiv* 683417. (doi:10.1101/683417)
108. Kinsella CM *et al.* 2019 Programmed DNA elimination of germline development genes in songbirds. *Nat. Commun.* **10**, 5468. (doi:10.1038/s41467-019-13427-4)
109. Wheeler DL *et al.* 2003 Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **31**, 28–33. (doi:10.1093/nar/gkg033)