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[2,3], Tri Haryoko[4], Knud A. Jønsson[5], Martin Irestedt[6], Qi Zhou[2,3,7], Patric Jern[8], Alexander Suh[1,9]

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

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

---

Main Text

Summary

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 >55% compared to the genome-wide density of <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 hypothesise that the avian W chromosome acts as a refugium for active ERVs, likely leading to female-biased mutational load that may influence female physiology similar to the "toxic-Y" effect in *Drosophila*. 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.

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 (PAR) 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 the first evidence that the avian W chromosome is not merely a graveyard of repetitive elements but a refugium of potentially active TEs that likely have sex-specific implications. Bird genomes are known to be repeat-poor with a mean TE content of <10% [19], but the first female assemblies based on short [20] or long reads [21–23] showed that the non-recombining W chromosome is >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, kakapo, paradise crow, zebra finch with heteromorphic sex chromosomes), we demonstrate that the avian W has generally accumulated ERVs and likely contains active ERVs as indicated by signatures of transcription and translation of W-linked ERVs. We therefore hypothesise 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.

Results and discussion

Enrichment of ERVs on the W chromosome

We analysed six avian genomes spanning the avian Tree of Life (**Figure 1A**) and representing the current standard for reference-quality genome assemblies [23,24]. Autosomes had between 7 and 11% TEs on average (**Figure 1B**, **Supplementary Table S2**, **Supplementary File S1**) and the Z chromosome had similar

or slightly more TE densities (4-17%), while the W chromosome stood out as having ~13-80% TEs (**Supplementary Table S2**). Noteworthily, we also found that the homomorphic W chromosome of emu is richer in TEs than the autosomes and Z (14% vs. 7.7% and 5%). 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 1B**, **Supplementary Table S2**). While long interspersed elements (LINEs) from the Chicken Repeat 1 (CR1) superfamily were the dominant repeats on autosomes and Z (cf. [19,25]), ERVs were the major component of the W chromosome and accounted for more than 50% of the assembled chromosome itself (**Supplementary Table S2**).

ERVs are Long Terminal Repeat (LTR) retrotransposons deriving from germline-inherited retrovirus integrations and exist mainly in two genomic forms [26,27]: 1) full-length elements with two long terminal repeats (likewise called LTRs) flanking its protein-coding genes necessary for retrotransposition; 2) 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 [28–30], we annotated full-length ERVs (fl-ERVs; **Supplementary Files S2-3**) and detected a large proportion of fl-ERVs on the W chromosome compared to the rest of the genome (**Figure 1C**, **Table 1, Supplementary Tables S3-12**). Despite the fact that the W chromosome accounted only for the 1-3% of the total length of assembled chromosomes (**Figure 1C**, **Supplementary 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 1D**, **Table 1**). The distribution of fl-ERVs deviated significantly (χ^2 test, p-values < 0.01) from a random distribution across all chromosomes (**Supplementary Table S4**), with an impoverishment of total ERV-derived bp on the autosomes (0-0.5 times fewer bp than expected) and an extreme accumulation on the W (6-56 times more bp than expected; **Supplementary Table S12**).

We propose a "refugium index" (formula **1.1**) to quantify the excess accumulation of TE-derived bp on an SLC compared to the rest of the genome. Since 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 vs. homogametic sex through the presence of an SLC (formula **1.2**). 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 transposition and genome instability, together likely 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 28 to 83% more fl-ERVs than males, and that even the emu has 13% more fl-ERVs in females than males despite largely homomorphic sex chromosomes [31,32]. 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 [8,23,24,33], 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 the 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.

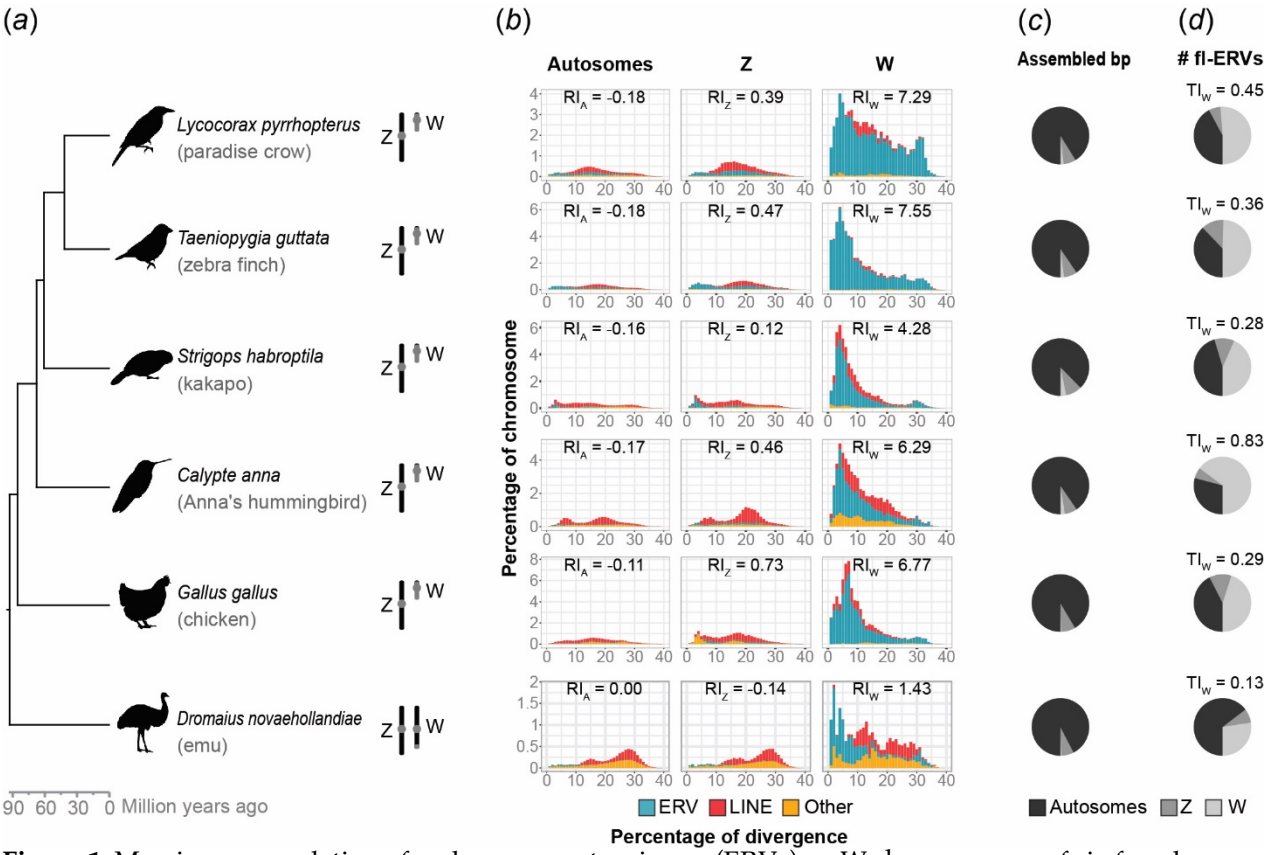

Figure 1. Massive accumulation of endogenous retroviruses (ERVs) on W chromosomes of six female reference-quality genome assemblies spanning the avian Tree of Life. *(a)* Avian time tree after [34] with schematic homomorphic or heteromorphic sex chromosomes. *(b)* TE landscapes of autosomes and sex chromosomes as stacked bar plots. Abundance of interspersed repeats (bp occupied) normalised 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 ERV numbers on autosomes and sex chromosomes as pie charts. The toxicity index (TI) of the W chromosome is reported for each species.

Transcription and translation of W-linked ERVs

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 (**Supplementary File S4**). 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. We then traced the presence of ERV proteins in male and female proteome data.

We analysed zebra finch, paradise crow, chicken, and emu for W-linked SNVs in genome re-sequencing and RNA-seq data. In each species, we found between 20 and 59 ERV subfamilies with W-linked SNVs (**Table 2**), with ERVL subfamilies being the most represented, and found evidence for the transcription of between 6 and 10 ERV subfamilies in female gonads (**Table 2**, **Supplementary Table S13**). For paradise crow, none of these SNVs were transcribed in female pectoral muscle, likely reflecting that somatic tissues generally show lower amounts of TE expression than gonads [35]. Our estimates of transcribed W-linked ERVs are likely just the tip of the iceberg, because we expect to identify W-linked SNVs only if those ERVs did 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).

Next, we analysed protein mass spectrometry data of white leghorn chicken gonads [36] with MaxQuant [37] for the presence of ERV proteins and found a higher quantity of these proteins expressed in females than in males as indicated by the high SILAC ratio H/L (**Supplementary File S5**). 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.

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 [38] and 28 genes on chicken W [21]). However, accumulating evidence shows that SLCs can have additional effects

[12,39,40]. For example, it is important to highlight that the Y-linked regulatory variation within populations of *Drosophila* can have genome-wide epistatic effects [14–16,41]. 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 [42]. 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 heterochromatin maintenance elsewhere relative to the SLC-lacking sex [13,42].

Recently, the Y chromosome repeat content has been linked to the destabilisation and loss of heterochromatin, which in turn is correlated to the shorter lifespan of the heterogametic sex [13,43]. 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 [13,44,45] whose expression is unleashed by heterochromatin loss. Possibly, the dysregulation of TEs due to heterochromatin loss is also associated to laminopathic diseases in *Drosophila* and humans [46]. 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 to be more accentuated (**Figure 2A** and **B**). The toxic-Y hypothesis has been recently investigated from a theoretical point of view in vertebrates with both XY and ZW systems [47] and put in contrast to the classic "unguarded-X" hypothesis [48–50], which proposes that the expression of recessive mutations on X/Z chromosomes is the cause of the shorter lifespan in the heterogametic sex. Sultanova et al. [47] 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 since 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.

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 hybridisation. This point may be not overly surprising in the context of Haldane's rule, which states that upon hybridisation, 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 [51]). 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 F2 hybrid embryo death [52]. 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 [53,54], 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 [55]. In addition, many TE repressor systems are in a sequence-specific arms race (e.g., piRNAs or KRAB-zinc finger proteins [53,56,57]), 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 likely sex-specific refugia of species-specific active TEs. Hybrid incompatibility due 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 gene disruption, genome instability [58], and hybrid dysgenesis [17]. TE/repressor mismatches can already occur during meiosis in the F1 hybrids, when recombination can separate the repressor from the controlled TEs (**Figure 2C** and **D**) [59]. 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), there are more chances for a mismatch between a repressor and intact TEs on the SLCs than for other chromosomes (**Figure 2D**).

For the birds analysed here, the W chromosome is likely 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 hybridisation in birds [60]. Thus, in addition to the preservation of dosage-sensitive genes [21,61], 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.

Conclusions

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 generalisable for all birds given our broad sampling of Palaeognathae, Galloanseres, 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. [62]), 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, similarly to the situation in *Drosophila* [17,52]. 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 [63]. 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 [64] or the attached-X/attached-X-Y karyotypes in *Drosophila* [65,66]) 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”.

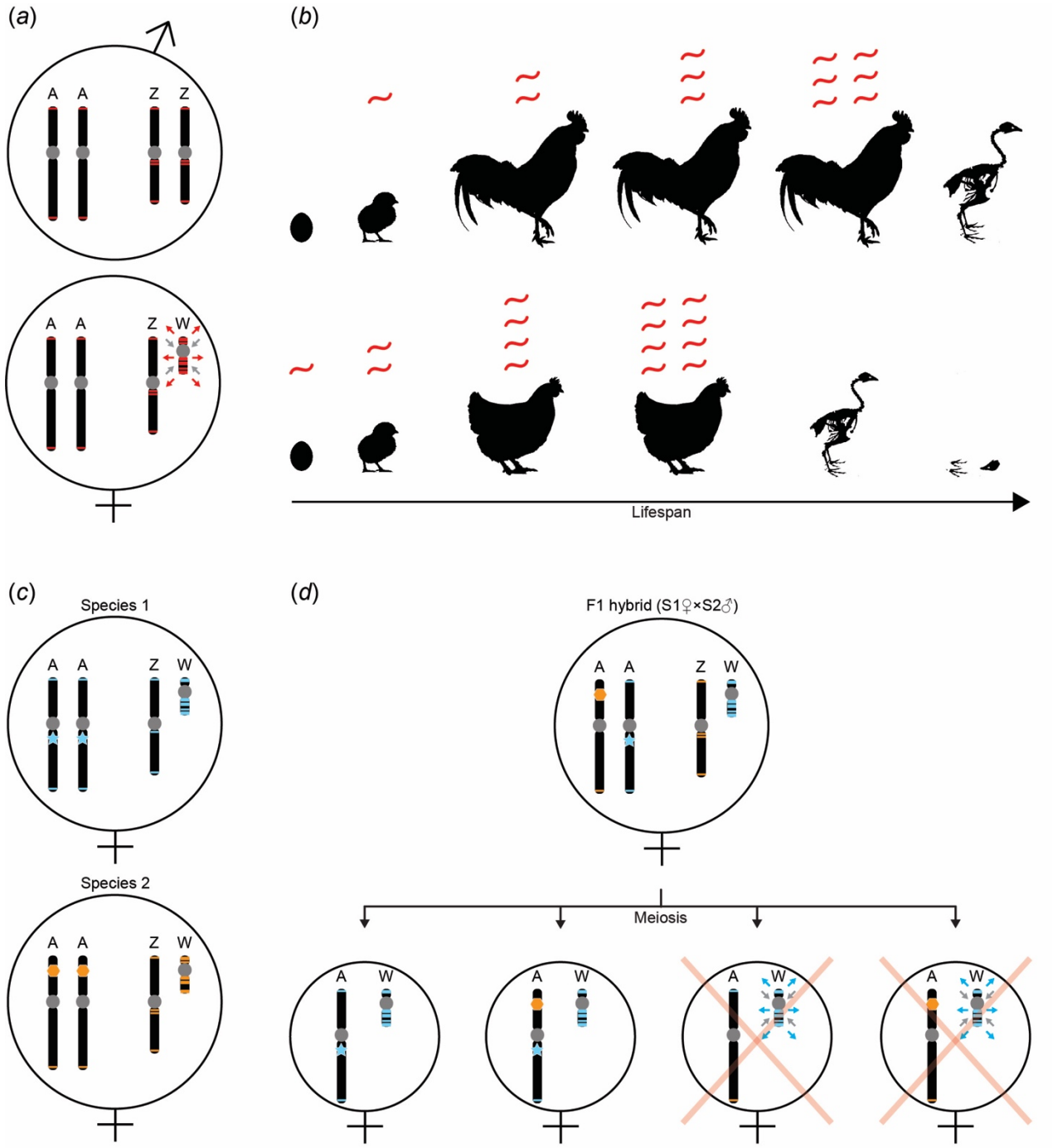

Figure 2. Synthesis of the consequences of the refugium hypothesis on micro- (*a* and *b*) and macroevolutionary (*c* and *d*) time scales. For simplicity, a schematic example of avian sex chromosomes is shown, but we expect these consequences for any ZW or XY system with SLC-linked intact transposable elements (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 due to the W chromosome in a female F1 hybrid between species 1 and species 2 of panel C. Schematic example of four possible meiotic products (oocytes) of the F1 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.

Materials and methods

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;* accession number pending) [23], emu (*Dromaius novaehollandiae*) [67], Anna's hummingbird (Calypte anna; GCA_003957555.2; bCalAnn1_v1.p) [24], kakapo (Strigops habroptila; GCA_004027225.2; bStrHab1.2.pri) [24] and zebra finch (*Taeniopygia guttata;* GCA_009859065.2; bTaeGut2.pri.v2) [24]. 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 [68] (originally uploaded on NCBI as of undetermined sex) and a female library of *Gallus gallus bankiva* (red junglefowl from Java) from [69]. The sexes of the individuals from [68] were determined using the SEXCMD with default sex markers [70]. Red junglefowl RNA-seq libraries of a female (ovary) and of a male (testes) were retrieved from [71]. 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 [69,72], RNA-seq libraries and protein mass-spectrometry libraries for five females (ovary) and five males (testes) [36].

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 library 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 male zebra finches from [73], and one RNA-seq library of testes and two of ovary from [74,75].

More details and accession numbers for all the libraries and genomic sequences here utilised can be found in **Supplementary Table S1**.

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 [23,76,77] while species-specific repeat libraries did not exist for emu, Anna's hummingbird and kakapo. Therefore, we *de-novo* characterised repetitive elements in these last three species using RepeatModeler2 [78]. We then concatenated the new *de-novo* libraries with the avian consensus sequences from Repbase [79], hooded crow [80], blue-capped cordon blue [81], flycatcher [82], and paradise crow [23], and used this final library to mask all six genomes with RepeatMasker [83].

Full-length ERV detection and abundance

Because LTR/ERV elements were the major TE group with strongest enrichment on W (see refugium index below), we decided to specifically quantify intact elements for these. Here we define full-length ERVs as possible retrotransposition-competent elements with relatively complete structures and potential to produce transcripts. In order to detect and quantify the presence of fl-ERVs, we used RetroTector [28] as well as LTRharvest [29] together with LTRdigest [30] on all the six avian genomes. RetroTector results were filtered for score >300 and presence of 5'- and 3'-LTR, as well as open reading frames (ORFs) with complete or partly complete *gag, pol* and *env* genes as previously described in [28,84]. LTRharvest results were filtered for false positive using LTRdigest in combination with HMM profiles of LTR retrotransposon-related proteins downloaded from Pfam [85] and GyDB [86].

Identification of SNVs of W-linked ERVs 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 analyzed in *Gallus gallus*, *Lycocorax pyrrhopterus* and *Taeniopygia guttata* (**Supplementary Table S1**). ERV translation was analyzed in *Gallus gallus* white leghorn breed [36]. RNA-seq and proteome libraries selected for this analysis were from gonad tissue with the exception of *Lycocorax pyrrhopterus* for which the RNA-seq data was generated from female pectoral muscle.

To identify W-linked SNVs from male/female read mapping, we used the WhatGene pipeline developed by [87] for SNV analyses of B chromosomes and germline-restricted chromosomes [88] 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.

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 [89]. The peptide

sequences were used as query database for MaxQuant 1.6.17.0 [37]. We used the experimental parameters described in [36] (**Supplementary File S5**); search results were filtered with a false discovery rate of 0.01. Second peptides, dependent peptides and match between runs parameters were enabled.

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 chi-square 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 (**Supplementary Tables S3-4**). Next, we calculated the refugium and toxicity indexes, which are described below for SLCs in general.

The refugium index (formula **1.1**) 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.

$$\textbf{1.1}\ \text{Refugium index} = \frac{\%TE_{obs} - \%TE_{exp}}{\%TE_{exp}}$$

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 (formula **1.2**) 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 x autosomes + 2 x Z or X) and the heterogametic sex (2 x autosomes + 1 x Z or X + 1 x 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 >95% complete ORFs of DNA transposons (i.e., transposase) and LINEs (i.e., ORF1 or ORF2) in tBLASTx or similar homology-based searches.

$$\textbf{1.2}\ \text{Toxicity index} = \frac{2n_{het} - 2n_{hom}}{2n_{hom}}$$

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.

Acknowledgments

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 lab and Johannesson lab for helpful discussions. We thank Erich Jarvis and the Vertebrate Genomes Project (VGP) for making their zebra finch, kakapo, and hummingbird assemblies available prior to publication. We thank Marco Ricci, Ivar Westberg, Jesper Boman and Diem Nguyen for their comments on the manuscript. This research was supported by grants from the Swedish Research Council Formas (2017-01597 to AS; 2018-01008 to PJ), the Swedish Research Council Vetenskapsrådet (2016-05139 to AS; 2018-03017 to PJ; 621-2014-5113 and 2019-03900 to MI), the SciLifeLab Swedish Biodiversity Program (2015-R14 to AS), Villum Foundation (Young Investigator Programme, project no. 15560 to KAJ), and from the Carlsberg Foundation (Distinguished Associate Professor Fellowship, project no. CF17-0248 to KAJ). 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. KAJ 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.

Tables

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

Species	Autosomes		Chromosome Z		Chromosome W		Males (2n)		Females (2n)		W toxicity index (LTRhd)
	RT	LTRhd	RT	LTRhd	RT	LTRhd	RT	LTRhd	RT	LTRhd	

Calypte anna	171	302	71	63	503	671	484	730	1490	1338	0,83
Dromaius novaehollandiae	129	261	0	32	1	111	258	586	260	665	0,13
Gallus gallus	392	232	74	68	244	246	932	600	1420	778	0,30
Lycocorax pyrrhopterus	396	558	83	88	439	678	958	1292	1836	1882	0,46
Strigops habroptila	354	911	102	225	458	871	912	2272	1828	2918	0,28
Taeniopygia guttata	298	552	149	184	450	723	894	1472	1794	2011	0,37

Table 2. Number of female-specific and thus W-linked SNVs detected at the genomic and transcriptomic levels for each species divided by ERV subfamily.

Species	# W-linked SNVs	# ERV subfamilies with W-linked SNVs	# transcribed SNVs	# transcribed ERV subfamilies
Dromaius novaehollandiae	178	44	6	6
Gallus gallus (red junglefowl)	1854	46	243	10
Gallus gallus (white leghorn)	681	59	8	6
Lycocorax pyrrhopterus	28	20	0	0
Taeniopygia guttata	134	57	36	1

References

1. Bachtrog D *et al.* 2014 Sex Determination: Why So Many Ways of Doing It? *PLOS Biol.* **12**, e1001899.
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 (N. Y).* **41**, 911–914. (doi:10.2307/2408899)
5. Mank JE. 2012 Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. *Chromosom. 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:https://doi.org/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:https://doi.org/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:https://doi.org/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 LP – 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.* **107**, 15826 LP – 15831. (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 LP – 833.
18. Bachtrog D. 2020 The Y Chromosome as a Battleground for Intragenomic Conflict. *Trends Genet.* **36**, 510–522. (doi:https://doi.org/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 LP – 117. (doi:10.1534/g3.116.035923)
23. Peona V *et al.* 2019 Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *bioRxiv* , 2019.12.19.882399. (doi:10.1101/2019.12.19.882399)
24. Rhie A *et al.* 2020 Towards complete and error-free genome assemblies of all vertebrate species. *bioRxiv* , 2020.05.22.110833. (doi:10.1101/2020.05.22.110833)
25. Zhang G *et al.* 2014 Comparative genomics reveals insights into avian genome evolution and adaptation. *Science (80-.).* **346**, 1311 LP – 1320. (doi:10.1126/science.1251385)
26. Kent T V, Uzunović J, Wright SI. 2017 Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160458. (doi:10.1098/rstb.2016.0458)
27. 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)
28. 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)
29. Ellinghaus D, Kurtz S, Willhoeft U. 2008 LTRharvest, an efficient and flexible software

for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18. (doi:10.1186/1471-2105-9-18)
30. 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)
31. 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.* **95**, 4415 LP – 4418. (doi:10.1073/pnas.95.8.4415)
32. 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)
33. 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)
34. 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)
35. Pasquesi GIM, Perry BW, Vandewege MW, Ruggiero RP, Schield DR, Castoe TA. 2020 Vertebrate Lineages Exhibit Diverse Patterns of Transposable Element Regulation and Expression across Tissues. *Genome Biol. Evol.* **12**, 506–521. (doi:10.1093/gbe/evaa068)
36. 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)
37. 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)
38. 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)
39. Chippindale AK, Rice WR. 2001 Y chromosome polymorphism is a strong determinant of male fitness in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5677–5682. (doi:10.1073/pnas.101456898)
40. 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)
41. 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)
42. 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)
43. 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)
44. Wei KH-C, Gibilisco L, Bachtrog D. 2020 Epigenetic conflict on a degenerating Y chromosome increases mutational burden in Drosophila males. *bioRxiv* , 2020.07.19.210948. (doi:10.1101/2020.07.19.210948)
45. 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)
46. 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:https://doi.org/10.1016/j.arr.2019.100995)
47. 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)
48. Trivers R. 1985 *Social evolution*.

49. 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)
50. 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 (N. Y).* **72**, 568–577. (doi:10.1111/evo.13434)
51. Serrato-Capuchina A, Matute DR. 2018 The Role of Transposable Elements in Speciation. *Genes (Basel).* **9**, 254. (doi:10.3390/genes9050254)
52. Ferree PM, Barbash DA. 2009 Species-Specific Heterochromatin Prevents Mitotic Chromosome Segregation to Cause Hybrid Lethality in Drosophila. *PLOS Biol.* **7**, e1000234.
53. 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:https://doi.org/10.1016/j.tig.2020.04.007)
54. Choi JY, Lee YCG. 2020 Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLOS Genet.* **16**, e1008872.
55. Huang CRL, Burns KH, Boeke JD. 2012 Active transposition in genomes. *Annu. Rev. Genet.* **46**, 651–675. (doi:10.1146/annurev-genet-110711-155616)
56. 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:https://doi.org/10.1016/j.tig.2017.08.006)
57. 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)
58. 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:https://doi.org/10.1016/j.gde.2017.07.014)
59. 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)
60. Petit RJ, Excoffier L. 2009 Gene flow and species delimitation. *Trends Ecol. Evol.* **24**, 386–393. (doi:https://doi.org/10.1016/j.tree.2009.02.011)
61. Bellott DW, Page DC. 2020 Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *bioRxiv* , 2020.07.09.196279. (doi:10.1101/2020.07.09.196279)
62. Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018 Why Do Sex Chromosomes Stop Recombining? *Trends Genet.* **34**, 492–503. (doi:https://doi.org/10.1016/j.tig.2018.04.001)
63. Allen R *et al.* 2020 A mitochondrial genetic divergence proxy predicts the reproductive compatibility of mammalian hybrids. *Proc. R. Soc. B Biol. Sci.* **287**, 20200690. (doi:10.1098/rspb.2020.0690)
64. 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)
65. Morgan L V. 1925 Polyploidy in Drosophila melanogaster with Two Attached X Chromosomes. *Genetics* **10**, 148–178.
66. Green MM, Piergentili R. 2000 On the origin of metacentric, attached-X (A-X) chromosomes in Drosophila melanogaster males. *Proc. Natl. Acad. Sci.* **97**, 14484 LP – 14487. (doi:10.1073/pnas.250483497)
67. Liu J *et al.* In prep. A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes.
68. 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)
69. 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:https://doi.org/10.1016/j.ygeno.2019.10.004)
70. 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.
71. McCarthy FM *et al.* 2019 Chickspress: a resource for chicken gene expression. *Database* **2019**. (doi:10.1093/database/baz058)
72. 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. *Zoolog. Sci.* **33**, 73–77. (doi:10.2108/zs150071)
73. Singhal S *et al.* 2015 Stable recombination hotspots in birds. *Science (80-.).* **350**, 928 LP – 932. (doi:10.1126/science.aad0843)
74. 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:https://doi.org/10.1016/j.cub.2018.03.067)
75. 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)
76. Wicker T *et al.* 2005 The repetitive landscape of the chicken genome. *Genome Res.* **15**, 126–136. (doi:10.1101/gr.2438004)
77. Warren WC *et al.* 2010 The genome of a songbird. *Nature* **464**, 757–762. (doi:10.1038/nature08819)
78. 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.* **117**, 9451 LP – 9457. (doi:10.1073/pnas.1921046117)
79. 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)
80. 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)
81. 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**. (doi:10.3390/genes10040301)
82. 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)
83. Smit AFA, Hubley R, Green P. 2015 RepeatMasker Open-4.0. See http://www.repeatmasker.org.
84. Hayward A, Cornwallis CK, Jern P. 2015 Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc. Natl. Acad. Sci.* **112**, 464 LP – 469. (doi:10.1073/pnas.1414980112)
85. El-Gebali S *et al.* 2018 The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432. (doi:10.1093/nar/gky995)
86. 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)
87. 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)
88. 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)
89. Wheeler DL *et al.* 2003 Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **31**, 28–33. (doi:10.1093/nar/gkg033)

Data Accessibility

All the data is publicly available and the code is accessible at http://github.com/ValentinaBOP/Wrefugium. All newly generated data was deposited in Sequence Read Archive (accession numbers XXXXXXXX-XXXXXXXX).

Authors' Contributions

VP and AS designed the study. VP analysed the data and wrote the first manuscript draft. VP and AS wrote and revised the subsequent drafts with input from all authors. KAJ and TH provided paradise crow samples. MI extracted paradise crow DNA. OMPG extracted paradise crow RNA. JB helped with repeat annotation. PJ ran RetroTector analyses. QZ and JL provided the emu genome assembly. AS supervised the study. All authors read and approved the manuscript.

Competing Interests

We have no competing interests.