The rattlesnake W chromosome: A GC-rich retroelement refugium with retained gene function across ancient evolutionary strata

Drew R. Schield[1], Blair W. Perry[2,3], Daren C. Card[4,5], Giulia I.M. Pasquesi[6], Aundrea K. Westfall[2], Stephen P. Mackessy[7], and Todd A. Castoe[2, §]

1. Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

2. Department of Biology, University of Texas at Arlington, Arlington, TX, USA

3. School of Biological Sciences, Washington State University, Pullman, WA, USA

4. Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

5. Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA

6. Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO, USA

7. School of Biological Sciences, University of Northern Colorado, Greeley, CO, USA

[§]To whom correspondence should be addressed: Todd A. Castoe, Department of Biology, University of Texas at Arlington, Arlington, TX 76010 USA. *Email:* todd.castoe@uta.edu *Phone:* 817-272-9084

Running title: Evolution and composition of snake W chromosomes

Keywords: *Crotalus*, recombination suppression, snakes, transposable elements, sex chromosomes, ZW sex determination

Abstract

Sex chromosomes diverge after the establishment of recombination suppression, resulting in differential sex-linkage of genes involved in genetic sex determination and dimorphic traits. This process produces systems of male or female heterogamety wherein the Y and W chromosomes are only present in one sex and are often highly degenerated. Sex-limited chromosomes (e.g., Y and W) contain valuable information about the evolutionary transition from autosomes to sex chromosomes, yet detailed characterizations of the structure, composition, and gene content of sex-limited chromosomes is lacking for many species. In this study, we characterize the female-specific W chromosome of the prairie rattlesnake (*Crotalus viridis*) and evaluate how recombination suppression and other processes have shaped sex chromosome evolution in ZW snakes. Our analyses indicate that the rattlesnake W chromosome is over 80% repetitive and that an abundance of GC-rich mdg4 elements has driven an overall high degree of GC-richness despite a lack of recombination. The W chromosome is also highly enriched for repeat sequences derived from endogenous retroviruses and likely acts as a 'refugium' for these and other retroelements. We annotated 219 putatively functional W-linked genes across at least two evolutionary strata identified based on estimates of sequence divergence between Z and W gametologs. The youngest of these strata is relatively gene-rich, however gene expression across strata suggests retained gene function amidst a greater degree of degeneration following ancient recombination suppression. Functional annotation of W-linked genes indicates a specialization of the W chromosome for reproductive and developmental function since recombination suppression from the Z chromosome.

Significance Statement: We report the first detailed analysis of the female-specific W chromosome in a snake. Our findings highlight distinctive features of W chromosome structure and function that provide expanded perspectives on transposable element evolution and the survival of genes related to female fitness.

Introduction

Sex chromosomes evolve from ancestral autosomes following the establishment of sex-determining genes and the onset of recombination suppression between the newly sex-linked genomic regions (Ohno 1967; Charlesworth 1996; Bergero and Charlesworth 2009). Over time, the absence of recombination between the sex chromosomes, i.e., Z and W or X and Y chromosomes in female- and male-heterogametic species, respectively, often leads to an accumulation of deleterious mutations and degeneration of the sex-limited chromosomes (Charlesworth and Charlesworth 2000; Charlesworth et al. 2005; Wright et al. 2016; Abbott et al. 2017). Recombination suppression does not necessarily occur simultaneously across the entire length of the newly formed sex chromosomes (Lahn and Page 1999; Handley et al. 2004; Charlesworth et al. 2005; Matsubara et al. 2006; Nam and Ellegren 2008; Vicoso et al. 2013). Instead, chromosomal regions can experience recombination suppression at various stages during the evolutionary transition from autosomes to sex chromosomes, leading to the establishment of 'evolutionary strata' that bear the signatures of this step-wise transition.

Low effective population size relative to autosomes and the lack of recombination renders selection against deleterious mutations inefficient on sex-limited chromosomes (e.g., Y or W), often resulting in an accumulation of repetitive content (Graves 2006; Bachtrog 2013). The accumulation of repeats, associated decay and loss of coding genes, and the high degree of heterochromatism on sex-limited chromosomes has promoted the view that they are genetic 'wastelands' that lack major functional roles outside of sex-determination (Bachtrog 2020). However, accumulating evidence from model systems indicates that sex-limited chromosomes may play prominent additional roles in shaping sex-specific traits (Piergentili 2010; Larson et al. 2018) and have broad effects on heterochromatin state and gene regulation across the genome (Jiang et al. 2010; Brown et al. 2020a; Nguyen and Bachtrog 2021). Potential deleterious regulatory effects of sex-limited chromosome degeneration led to the 'toxic-Y' hypothesis in *Drosophila* (Marais et al. 2018; Brown et al. 2020b; Wei et al. 2020), whereby the high Y-linked mutational load may correlate with the decay of physiological function and shorter lifespans in males (Sultanova et al. 2020). As a major component of this mutational load, studies have observed a disproportionate abundance of full-length and potentially transpositionally-active transposable elements (TEs) present on the sex-limited chromosomes of both XY and ZW systems (Steinemann and Steinemann 1992; Bachtrog 2003; Peona et al. 2021). Indeed, the presence of potentially active retroviral-like elements (e.g., endogenous retroviruses; ERVs) in bird species with different degrees of W chromosome degeneration spurred the development of the 'refugium hypothesis' (Peona et al. 2021), which predicts that the female-specific W chromosome, and male-specific Y chromosome in XY species, represent

refugia for retroelements that may have disruptive impacts on heterochromatin formation and gene regulation.

Even in the midst of substantial genetic decay, sex-limited chromosomes may sometimes retain or acquire protein-coding genes, and in some cases may experience an amplification of specific genes through duplication events (Mank et al. 2014). These genes may be dosage-sensitive, underlie sex-specific functions, and/or be involved in meiotic drives and other forms of sexual conflict (Meiklejohn and Tao 2010; Parsch and Ellegren 2013). An associated prediction is that duplications evolve to compensate for lower overall transcription across heterochromatic sex-limited regions or because they confer sex-specific fitness advantages (Bachtrog 2020). For example, the neo-Y chromosome of *Drosophila miranda* has experienced a massive accumulation of genes that are dosage-sensitive or involved in testis-specific functions via duplication events (Bachtrog et al. 2019; Ellison and Bachtrog 2019). In mice, the Y chromosome is gene-rich due in large part to a parallel accumulation of genes that distort X chromosome transmission and restore equal population sex ratios when amplified on the Y chromosome, and which are necessary for male fertility (Soh et al. 2014; Kruger et al. 2019). Similarly, the avian W chromosome has experienced an amplification of the *HINTW* gene hypothesized to play roles in oogenesis and female fecundity (Ceplitis and Ellegren 2004). In addition to the amplification of gene products, gene duplication on sex-limited chromosomes can facilitate gene conversion events (Backström et al. 2005), presenting an evolutionary solution to guard against deleterious mutation accumulation impacting critical genes.

Among amniotes, snakes are a valuable model system for studying sex chromosome evolution, as male and female heterogamety has evolved independently in different lineages (Fig. 1A; Matsubara et al. 2006, 2019; Gamble et al. 2017; Augstenová et al. 2018), providing useful comparisons to established XY and ZW model systems. Caenophidian snakes *sensu* Zaher et al. (2019), including the Acrochordoidea and Colubriformes (colubrids, elapids, viperids, and relatives), possess homologous ZW chromosomes (Rovatsos et al. 2018; Fig. 1A). Analyses of this group are therefore of particular interest because they can provide insight into the evolutionary consequences of female heterogamety in a sex chromosome system that evolved independently of ZW chromosomes in birds (i.e., avian and caenophidian ZW chromosomes were derived from different ancestral autosomes; Matsubara et al. 2006; O'Meally et al. 2012). Despite cytogenetic studies on snake sex chromosomes beginning decades ago (e.g., Becak et al. 1964; Ohno 1967), far less is known about snake sex chromosome structure, function, and evolution compared to mammal and bird systems. Previous studies have demonstrated variation in the degree of divergence between Z and W chromosomes among caenophidian snake lineages (Matsubara et al. 2006), the presence of multiple evolutionary strata on the Z chromosome, and a lack of global dosage

compensation in females (Vicoso et al. 2013; Yin et al. 2016; Schield et al. 2019). One of the hypothesized evolutionary strata on the Z chromosome exhibits multiple signatures of relatively recent recombination suppression (i.e., the so-called 'recent stratum' described in Schield et al. 2019), including a lesser degree of degeneration than older strata and high levels of female-specific heterozygosity. The recent stratum may be unique to pitvipers (it was detected in *Deinagkistrodon*, *Sistrurus*, and *Crotalus* species), although additional comparative analyses and divergence time estimates to confirm the absence of recombination suppression in this region in other lineages have not been conducted. Additionally, while a previous study (Yin et al., 2016) estimated the oldest stratum to be at least 66.9 million years old, this estimate was based on a limited sampling of ZW gametologs, and the ages of more recent recombination suppression events remain unknown.

Despite progress, our understanding of the structure and composition of the snake W chromosome remains far from complete, and represents a major gap in our understanding of sex chromosome evolution across ZW systems in vertebrates. Previous investigations have been limited due primarily to the difficulty in reconstructing the W chromosome, which is the result of the heterochromatic and repetitive complexity, along with hemizygosity of the chromosome in females leading to lower genome sequencing coverage. This has in turn limited inferences regarding the number and age of evolutionary strata, repeat and gene content, and whether phenomena observed in other systems (e.g., the toxic-Y and refugium hypotheses) are also relevant in snake sex chromosome evolution. However, developments in sequencing technologies, such as linked-read sequencing (Weisenfeld et al. 2017), have improved the ability to assemble scaffolds spanning repeat-rich genomic regions. In this study, we use this technology to reconstruct the prairie rattlesnake (*Crotalus viridis*) W chromosome, and use these data to complement previously assembled genome data for autosomes and the Z chromosome of this species (Schield et al. 2019). We use these assemblies to characterize gene and repeat content on sex chromosomes to address several outstanding or otherwise poorly-resolved questions in snake sex chromosome evolution: 1) how many evolutionary strata are present between Z and W chromosomes and how long ago did recombination suppression occur, 2) what is the repeat and nucleotide composition of the W chromosome, 3) does the W chromosome represent a refugium for retroviral-like elements, as observed in other taxa, 4) do W-linked genes show evidence of retained function across evolutionary strata, and 5) is there evidence of W-specific gene duplications that may be relevant to female-specific traits or fitness?

Results

Evolutionary strata on the rattlesnake sex chromosomes

We identified a total of 23.7 Mb of W chromosome sequence from an assembly of a female prairie rattlesnake, represented by 2,027 scaffolds (Table 1; Supplementary Tables S1-S2) assembled using high-quality 10x Genomics linked-reads (Supplementary Fig. S1). Of these, 1,113 scaffolds (13.9 Mb) could be confidently anchored to the Z chromosome based on sequence homology (Fig. 1B). The density of homologous regions is highest in the previously identified recent stratum and the centromere, matching sequence-based evidence of ZW homology in the recent stratum (Schield et al. 2019) and cytological evidence for the relative position of the viperid snake W centromere (Baker et al. 1972; Matsubara et al. 2006). There is also evidence of ZW homology across the length of the Z chromosome, albeit at lower levels, indicating more substantial W chromosome degeneration in these regions. We annotated 219 genes across these W chromosome scaffolds.

We calculated divergence between Z and W gametologs to investigate the presence of evolutionary strata and to date the onset of recombination suppression. Synonymous divergence, d_S, varies considerably (mean $d_S = 0.24 \pm 0.19$), with gametologs falling into three spatial clusters along the Z chromosome (Fig. 1C). While there is considerable variation in d_S estimates for gametologs within clusters, broad differences in d_S among clusters suggest differential timing of recombination suppression between Z and W chromosomes across these regions (Fig. 1D; Supplementary Table S3). We used these differences to infer the approximate locations of at least two evolutionary strata between the Z and W chromosomes, supporting previous hypotheses based on studies in other viperid species (Vicoso et al. 2013; Yin et al. 2016; Schield et al. 2019). Genes in clusters 1 and 2 correspond to older strata across the majority of the Z chromosome (and there may be multiple undetectable strata across this broad region), with the highest d_S values on average (mean $d_S = 0.33 \pm 0.2$). Cluster 3 corresponds to the recent stratum and contains genes with significantly lower d_S than the older strata (mean $d_S = 0.16 \pm 0.12$; Mann-Whitney ∪ test; p-value < 0.0001), although we note that d_S distributions between strata do not have equal variance (Levene's test, $p < 2.2 \times 10^{-16}$). Using lineage-specific mutation rate estimates derived from synonymous and four-fold degenerate sites (see Methods), we transformed mean d_S values to divergence times to determine an approximate date range of recombination suppression in the older and recent strata. Based on this, we estimate that recombination suppression occurred approximately 87-102 MYA in the older strata and 40-47 MYA in the recent stratum (Fig. 1E).

Comparative analyses using mapped reads across the Z chromosome from female and male garter snakes (a distantly related species with ZW sex chromosomes) indicates relative female coverage (log_2FM) values consistently near -1 ($log_2FM = -0.81 \pm 0.54$) across the recent stratum (Fig. 2). In contrast, each of the pitviper species analyzed has log_2FM values in the recent stratum that are intermediate between older strata and the PAR (Fig. 2). These results suggest that recombination suppression in this region occurred independently in the garter snake and pitviper lineages since their divergence, consistent with the inferred age of the pitviper recent stratum being younger than the divergence time between the garter snake and pitvipers (61.7 MYA; timetree.org; Kumar et al. 2017).

The W chromosome is GC- and repeat-rich

Our comparative analyses indicate that the rattlesnake W chromosome has a higher GC content (mean GC = 43.9 ± 0.041%) than the Z chromosome (36.8 ± 0.041%) and autosomes (36.3 ± 0.04%; Mann-Whitney ∪ tests, *p*-values < 2.2×10^{-16}; Fig. 3A). The W chromosome also has a higher proportion of CpG sites than both the Z chromosome and autosomes (W = 4.8 ± 0.9%, Z = 0.97 ± 0.51%, autosomes = 0.96 ± 0.54%; Mann-Whitney ∪ tests, *p*-values < 2.2×10^{-16}; Fig. 3B). These results are surprising because the W chromosome is non-recombining, meaning that an accumulation of GC content though GC-biased gene conversion is not a plausible explanation for the higher overall GC content on the W chromosome relative to recombining chromosomes. To explore this result in the context of other amniotes with degenerated sex-limited chromosomes, we measured GC and CpG content in chicken, zebra finch, human, and mouse. In these comparisons, GC content on sex-limited W and Y chromosomes is significantly lower than autosomes and Z(X) chromosomes (Supplementary Table 4; Mann-Whitney ∪ tests; *p*-values < 0.05), the opposite pattern observed in prairie rattlesnake (Fig. 3A). Similarly, CG dinucleotide (i.e., CpG) content is lower on the W and Y chromosomes of the bird and mammal species with the exception of the zebra finch (Mann-Whitney ∪ tests, *p*-values < 2.2×10^{-16}; Fig. 3B), which has slightly higher W-linked CpG content. Compared to the other species, CpG content on the rattlesnake W chromosome stands out as being nearly five-fold higher than the Z chromosome and autosomes (Fig. 3B).

The rattlesnake W chromosome also contains very high proportions of repeat element-derived sequences (Fig. 3C), consisting of 81.15% total repetitive content (more than two-fold higher than rattlesnake autosomes; Fig. 3D) derived from a diversity of TEs. Such high TE density in the rattlesnake W is similar to sex-limited chromosomes of other species (Bachtrog 2005; Graves 2006; Smeds et al. 2015; Peona et al. 2021) and is consistent with the prediction that there is less efficient selection on the sex-limited chromosome to purge deleterious mutations due to the absence of recombination and lower effective population size than other chromosomes (Charlesworth and Charlesworth 2000). This finding strongly

contrasts with the 15.39% repeat content reported in Singchat et al. (2020) based on the previously reported Indian cobra (*Naja naja*) W chromosome scaffold (Super-Scaffold_1000010; Suryamohan et al. 2020). However, we show that this scaffold was incorrectly identified as the W chromosome, and instead represents an autosomal chromosome scaffold (see Discussion; Supplementary Appendix; Supplementary Fig. S2A-B). Further, using the same approaches we used to identify candidate W-linked scaffolds in the rattlesnake, we identified a revised (and distinct) set of putative W chromosome scaffolds in the female Indian cobra assembly (Supplementary Fig. S2C; Supplementary Data) totaling 35.9 Mb in length that do not include Super-Scaffold_1000010.

The most abundant TE family on the rattlesnake W chromosome are mdg4 (formerly known as 'gypsy') retroviral-like LTR retrotransposons, which constitute 23.5% of our assembled W chromosome sequence (Fig. 3E; Supplementary Table S5). Autosomes and the Z chromosome contain much lower proportions of mdg4 elements overall (2.23% and 6.64%, respectively), with lower mdg4 retroelement densities across autosomal and Z-linked regions (i.e., 10 kb sliding windows) compared to the W chromosome (Mann-Whitney ∪ tests, $p < 2.2 \times 10^{-16}$). Several potential mechanisms could explain this pattern. One possibility is that mdg4 elements on ancestral proto-sex chromosomes have persisted (i.e., experienced reduced decay) on the W chromosome following recombination suppression; a large proportion of relatively ancient mdg4 elements on the W chromosome would support this hypothesis. Alternatively, this pattern could be driven by an accumulation of mdg4 elements on the W chromosome independent of their evolution on the Z chromosome. The age distribution of mdg4 elements on the W chromosome (i.e., CpG-corrected Kimura distance) indicates prolonged activity, including a large proportion of recently active mdg4 elements (Fig. 3C). W-linked mdg4 retroelements also tend to be longer in length than those on the Z chromosome and autosomes (mean length = 625.56 bp on the W chromosome versus 475.8 bp on autosomes and the Z chromosome; Supplementary Fig. S3; Mann-Whitney ∪ tests, p-values $< 4 \times 10^{-11}$). It is also notable that in our analysis of full-length retroelements (below) we identify 327 full-length and potentially transpositionally-active mdg4 elements (i.e., with paired LTRs and open reading frames with intact *gag* and *pol* genes) on the W chromosome. The relative abundance of W-linked mdg4 elements potentially capable of self-replication argues for their accumulation after recombination suppression between the sex chromosomes.

mdg4 retroelements drive GC-richness on the W chromosome

Given the high frequency of W-linked mdg4 retroelements, we asked if these elements may substantially contribute to compositional patterns on the W chromosome, such as the observed GC-richness. Indeed, W-linked mdg4 retroelement regions are GC-rich (mean GC = 47.4 ± 5.1% SD), significantly more so

than L1 LINEs and CR1/L3 LINEs which are otherwise fairly abundant TE families in the entire rattlesnake genome and on the W specifically (34 ± 5.6% and 44.8 ± 7.8% GC, respectively; Mann-Whitney ∪ tests, p-values < 2.2×10^{-16}; Supplementary Table S5). Mdg4 elements are also significantly more GC-rich than the total background of non-mdg4 TEs (mean GC = 43.4 ± 12.8%; $p < 2.2 \times 10^{-16}$). We therefore tested whether regional GC content is associated with mdg4 element density, and find that GC content and mdg4 density are positively correlated across the W chromosome (Spearman's rank correlation coefficient $r = 0.46$, $p < 2.2 \times 10^{-16}$; Fig. 3F). This association persists when controlling for CpG content and total repeat content (Spearman's partial correlation coefficient $r = 0.15$, $p = 1.55 \times 10^{-11}$). Further, we find no correlation between GC content and total repeat content when mdg4 elements are removed ($r = 0.046$, $p = 0.18$; Fig. 3G). These results support a view that the W chromosome has evolved higher GC content than autosomes and the Z chromosome primarily due to the proliferation of GC-rich LTR retroelements.

The W chromosome is a refugium for retroelements

We tested the hypothesis that the rattlesnake W chromosome acts as a refugium for retroviral-like elements by examining the frequencies of TE families, including full-length LTRs, across the genome and testing whether these fit a uniform distribution or if the W chromosome is exceptional with regard to genome-wide TE distributions. We used a χ^2 framework (Peona et al. 2021) to compare observed and expected TE frequencies and to calculate a refugium index for autosomes, the Z, and the W, where positive RI values indicate an excess of elements compared to a uniform distribution. Indeed, we find that the W chromosome contains a significant excesses of total repeat elements (refugium index (RI) = 1.8), ERVs (RI = 16.64), mdg4 (RI = 7.5), and L1 LINE elements (RI = 4.7; Fig. 4A-D; Table 2; Supplementary Table S6). The observed distributions of TEs in each of these categories deviated significantly from a uniform expectation (total repeat $\chi^2 = 14{,}226{,}221$, $df = 2$, $p < 2.2 \times 10^{-16}$; ERV $\chi^2 = 8{,}724{,}789$, $df = 2$, $p < 2.2 \times 10^{-16}$; mdg4 $\chi^2 = 17{,}352{,}014$, $df = 2$, $p < 2.2 \times 10^{-16}$; L1 LINE $\chi^2 = 6{,}583{,}173$, $df = 2$, $p < 2.2 \times 10^{-16}$). The Z chromosome also has RI values consistent with an excess of TEs, yet these values are consistently lower than those observed on the W (Fig. 4A-D), with the exception of CR1 LINEs, for which the Z chromosome has a higher RI (0.059) than the W chromosome (0.028); observed frequencies of CR1 LINEs also deviate significantly from the uniform expectation ($\chi^2 = 14{,}663$, $df = 2$, $p < 2.2 \times 10^{-16}$). In contrast to patterns on the sex chromosomes, autosomes do not show evidence of excess TEs (i.e., RI < 0; Table 2).

We identified a total of 15,961 genome-wide full-length LTR retrotransposons (fl-LTRs), defined as having intact LTR and protein domains (see Methods), that were non-randomly distributed across the

genome ($\chi^2 = 384.21$, $df = 2$, $p < 2.2 \times 10^{-16}$), with significant enrichment on the sex chromosomes. Of these, 572 (3.6%) are located on the W chromosome, corresponding to an RI of 0.89. The Z chromosome contains 1,786 fl-LTRs (11.2%; RI = 0.29). Autosomes contain the remaining 13,603 fl-LTRs (85%), significantly fewer than expected under a uniform expectation and corresponding to a RI of -0.047. In addition to the W chromosome having the highest fl-LTR RI, these full length retroelements constitute 13.2% of total W-linked sequence. These results, along with evidence of high RI for specific TE families, indicate that the rattlesnake W chromosome acts as a refugium for potentially transpositionally-active LTR retroelements. We further tested for evidence of sex differences in LTR-derived 'toxicity' by calculating a toxicity index (Peona et al. 2021) based on identified fl-LTRs on autosomes and the sex chromosomes (see Methods). This analysis suggests no sex bias in toxicity, as the calculated index is very near zero (toxicity index = -0.03).

W-linked gene expression and female-specific function

Gene expression analyses using female and male RNAseq data identified evidence of expression for 145 (66.2%) of the 219 annotated genes on the W chromosome (Supplementary Table S7). Of these, 137 W gametologs (62%) show evidence of being expressed in female tissues (mean female TPM > 0), and 31 (14.2%) have female-biased expression based on formal tests of 'upregulation' compared to males (IHW *p*-values < 0.05; Fig. 5A; Supplementary Fig. S4). This analysis revealed only two other female-biased genes on other chromosomes, indicating disproportionate W-linkage of genes with female-biased expression. Additionally, a total of 103 genes show evidence of female expression (raw expression count > 0 in at least one female tissue) and no evidence of male expression, further emphasizing W-linkage and divergence between Z and W gametologs for these genes. Identified 1:1 ZW gametologs with evidence of female expression are similar in abundance in both the recent and older evolutionary strata (45 and 44 genes, respectively), although the older strata contain a higher proportion of genes with female-biased expression (25.4%) than the recent stratum (12.7%). Overall, evidence of W-linked gene expression indicates a degree of retained gene function across both recent and older evolutionary strata.

We tested whether genes on the W chromosome show any relevant links to female-specific traits and biological processes by characterizing functional annotations for W-linked genes using gene ontology (GO) term, pathway, and protein class classifications. Multiple enriched GO terms (i.e., FDR-corrected *p*-value < 0.05) are directly relevant to embryogenesis, including embryonic organ development, skeletal system morphogenesis, and embryonic skeletal system development (Fig. 5B). Other non-enriched terms represented by multiple W-linked genes relate to developmental processes, including facial development, cranial skeletal system development, hindbrain development, and nerve development. W-linked genes are

not significantly enriched for any specific pathways, however several pathways relevant to embryo development involve genes on the W chromosome, including: activation of HOX genes during differentiation, activation of anterior HOX genes in hindbrain development during early embryogenesis, and programmed cell death (Supplementary Fig. S5). Other functional categories involving W-linked genes relate to chromatin regulation (Fig. 5B), including: covalent chromatin modification (biological process; 9 genes), nuclear chromatin (cellular component; 8 genes), acetyltransferase complex (cellular component; 4 genes), and the histone acetyltransferases acetylate histones pathway (4 genes). Protein classifications indicate that the W chromosome also contains genes that function as gene-specific transcriptional regulators and protein modifying enzymes (Fig. 5C). One of these genes, *LMTK3*, is a positive regulator of the estrogen receptor *ESR1* (Giamas et al. 2011), which plays a putative role in dosage compensation on the Z chromosome in rattlesnakes (Schield et al. 2019) and female-biased gene expression in other reptiles (Rice et al. 2017).

Evolution of W-linked genes: survival, duplication, selection, and gene conversion

A large percentage of W-linked genes (49.6%) occur in the recent stratum, representing a significant enrichment over the proportion of Z-linked genes in this region (19.7%; Fisher's exact test, $p = 2 \times 10^{-7}$). In contrast, the older stratum of the W chromosome is depleted of genes relative to the Z chromosome (50.4% versus 80.3%, respectively; $p = 0.0037$). This suggests a greater degree of gene persistence/retention (or a slower decay rate) in the recent stratum of the W chromosome, though this pattern could be a simple consequence of more ancient recombination suppression in older strata. To account for this, we calculated the rate of W-linked gene decay in each stratum based on the number of ancestral genes lost on the W chromosome and estimated divergence times. Here, we calculate a W-linked recent stratum decay rate of 5×10^{-6} (i.e., five genes lost per million years), more than two-fold slower than the W-linked older stratum rate (1.2×10^{-5}), and consistent with a greater degree of gene retention in the recent stratum.

Translocation from autosomes is another mechanism to combat gene decay on the W chromosome (Mank 2012), in which autosomal genes are transferred to the W chromosome after recombination suppression between the sex chromosomes. We tested for evidence of translocated genes by identifying W-linked genes that lack a matched Z gametolog and are also not orthologous to genes on *Anolis* chromosome 6 (which shares common ancestry with caenophidian snake ZW sex chromosomes). We find 17 genes (7.7% of W-linked genes) with evidence of translocation to the W from autosomes (Supplementary Table S8), indicating that both ancestral gene survival and gene gains from autosomes contribute to gene content on the W chromosome. Functional annotation of translocated genes do not show statistical

enrichment for specific GO terms or pathways (Supplementary Fig. S6A-B), however several genes have known roles in immune function (*H2Q9* and *RBM14*; see below) and the gene *VWA5A* may function as a tumor suppressor in humans (Martin et al. 2003). Other candidate autosome-W-translocated genes are uncharacterized (Supplementary Table S8).

We further tested for evidence of W-specific gene duplications (i.e., ampliconic genes; Bachtrog 2013) in the rattlesnake by comparing sequencing depths between autosomes and the W chromosome, with the expectation that single-copy W gene depth would be half of that observed for autosomal genes. We infer that 21% of genes on the W chromosome are present in two or more copies, with a subset showing evidence of even greater degrees of amplification (Fig. 6A; Supplementary Table S9). Several of these genes lack apparent Z gametologs and are annotated as *MLV*-related proviral *ENV* polyprotein, *ERV3-1*, and *ERV-2*, which are present in at least 94, 47, and 6 total copies, respectively. Amplification of these genes is logical given that they are derived from retroviral-like elements (Cohen et al. 1985; Evans et al. 2003), which are abundant and active on the W chromosome (Fig. 4). Other ampliconic genes include *RBM14*, which is involved in the activation of the innate immune response in humans (Morchikh et al. 2017), and multiple members of the vertebrate histocompatibility complex (*H2Q9*, *RT1B*, and *HA1F*). We infer that *H2Q9* and *RBM14* were translocated from autosomes to the W chromosome, with evidence of W-specific duplications indicating subsequent amplification following translocation. While these results suggest that gene duplication may have produced an expanded immune gene complement in females, we note that our approach cannot confirm whether these duplicates are functional because evidence for the multi-copy nature of these genes is based only on relative read depths rather than multiple reconstructed gene models in our annotation (likely due to their repetitive and structural complexity). Inferred duplicated genes were not significantly enriched for specific GO terms or pathways (Supplementary Fig. S6C-D), though several genes are involved in response to estradiol, tissue and neuronal development, and skeletal morphogenesis.

Because it does not experience sexual recombination and has a low effective population size, we predict that the effects of linked selection and genetic drift hinder the efficiency of selection on the W chromosome to a greater extent than the rest of the genome. Comparing rates of between-species sequence divergence for the sex chromosomes provides a means to test whether these factors have indeed differentially influenced sequence evolution on the Z and W chromosomes since recombination suppression. We calculated the ratio of nonsynonymous to synonymous divergence (d_N/d_S) between prairie rattlesnake and five-pace viper for Z and W gametologs, and compared the distributions of d_N/d_S with the expectation that higher values are indicative of less efficient purifying selection (Charlesworth

and Charlesworth 2000). Importantly, we restricted these analyses only to genes for which we could identify reciprocal best hit orthologs to avoid comparisons between gene duplicates (see Methods). The W chromosome has significantly higher d_N/d_S (0.44 ± 0.29) than the Z chromosome (0.26 ± 0.22; Mann-Whitney U test, p-value = 2.8×10^{-6}; Fig. 6B; Supplementary Table S3), consistent with reduced efficacy of purifying selection on the non-recombining W chromosome. This result is consistent across evolutionary strata, with higher W-linked d_N/d_S in stratum 1 (i.e. older strata; W $d_N/d_S = 0.41 \pm 0.29$, Z $d_N/d_S = 0.26 \pm 0.26$; p-value = 0.0034) and stratum 2 (recent stratum; W $d_N/d_S = 0.46 \pm 0.3$, Z $d_N/d_S = 0.26 \pm 0.19$; p-value = 0.0003).

In addition to differences in effective population size between the sex chromosomes and the effects of selective interference on the W chromosome, it experiences a female-specific mutation rate because it is unique to the female germline. Investigations comparing rates of divergence between autosomes and the Z chromosome in snakes have demonstrated male-biased mutation rates (Vicoso et al. 2013; Schield et al. 2021), similar to patterns observed in birds (e.g., Axelsson et al. 2004; Smeds et al. 2015). We therefore expected to find higher mutation rates on the Z chromosome than W chromosome. Surprisingly, we find higher d_S between the prairie rattlesnake and five-pace viper W chromosomes (0.18 ± 0.21) than between the Z chromosomes (0.1 ± 0.09; p-value 0.02; Fig. 6C; Supplementary Table S3).

Faster divergence on the W chromosome could be driven by a higher female-specific mutation rate than previously estimated, although a more likely mechanism is intrachromosomal gene conversion, which has been observed among sex-limited chromosomes of diverse taxa (Slattery et al. 2000; Rozen et al. 2003; Backström et al. 2005; Wright et al. 2014; Chang and Larracuente 2019). A uniting feature in these examples is the presence of gene duplications, enabling gene conversion between paralogs on the sex-limited chromosome (Connallon and Clark 2010). The abundance of amplified gene copies on the rattlesnake W chromosome may present a similar repetitive substrate for intrachromosomal gene conversion. To test for gene conversion on the W chromosome, we compared the proportion of GC bases at third codon positions (GC3) between ZW gametologs, following Smeds et al. (2015). Here, the expectation under the GC-biased gene conversion model (Galtier et al. 2001) is that W-linked genes will have lower GC3 than their Z-linked counterparts in the absence of gene conversion on the W chromosome. However, the GC3 distribution for W gametologs ($51.2 \pm 12\%$) is not significantly lower than Z gametologs ($50 \pm 11\%$; Welch's two-sample t-test, p-value = 0.27; Fig. 6D; Supplementary Table S3), and 56% of W genes have higher GC3 than the matching Z gametolog. This finding contrasts with the avian W chromosome, for example, which contains few ampliconic genes with higher GC3 than Z gametologs (e.g., *HINTW*; Ceplitis and Ellegren 2004; Backström et al. 2005; Smeds et al. 2015; Rogers

et al. 2021), and suggests that intrachromosomal gene conversion may be comparatively frequent on the snake W chromosome, playing a role in shaping patterns of W-linked synonymous sequence divergence and GC content.

Discussion

Evolution of the distinctive rattlesnake W chromosome

The rattlesnake W chromosome possesses a number of distinctive structural and evolutionary features that provide an expanded perspective of how non-recombining sex chromosomes evolve in vertebrates, illustrating a broader diversity of evolutionary trajectories for sex-limited chromosome evolution in comparison to established models from mammals and birds. We infer that recombination suppression between the sex chromosomes occurred in at least two steps, with the initial formation of an ancient evolutionary stratum (or strata) roughly 87 million years ago followed by the formation of the recent stratum 40 million years ago (Fig. 1). The presence of at least two evolutionary strata is consistent with several previous investigations (Vicoso et al. 2013; Yin et al. 2016; Schield et al. 2019), although these previous studies, which lacked the resolution of the present W chromosome assembly, were unable to determine whether the recent stratum represents a unique recombination suppression event in pitvipers, or if this event took place earlier in colubriform or even caenophidian snake evolution (i.e., caenophidans possess homologous ZW sex chromosomes; Rovatsos et al. 2018). Based on comparative analyses of colubriform species and divergence between ZW gametologs, we infer that recombination suppression in the recent stratum occurred after the split between pitvipers and the garter snake (Figs. 1–2), suggesting that restricted recombination evolved independently in the homologous region of the sex chromosomes of other lineages (e.g., colubrids and elapids). Whether recombination suppression in the recent stratum occurred prior to the common ancestor of Viperidae is unclear. Our divergence time estimate is very close to the estimated split between true vipers and pitvipers (42.8 MYA; Zheng and Wiens 2016), and additional analysis of sex chromosomes in viperine species will be necessary to confirm whether the recent stratum is a shared feature of the Viperidae.

The majority of the caenophidian sex chromosome pair, however, experienced more ancient recombination suppression (Figs. 1–2). While we suspect that this region may encompass multiple evolutionary strata, recombination suppression likely occurred too closely in time to distinguish between independent events (Fig. 1C–D). The inferred age range of the older strata is coincident with the ancestral split between Caenophidia (including Colubriformes and Acrochordoidea) and Henophidia (including

boas and pythons) roughly 91 MYA (Kumar et al. 2017). This is particularly interesting because henophidian snakes have evolved independent XY and ZW sex chromosomes multiple times (Gamble et al. 2017; Augstenová et al. 2018), while all caenophidian snakes sampled to date have homologous ZW sex chromosomes (Matsubara et al. 2006; Rovatsos et al. 2015, 2018). The age of the older strata and ubiquitous presence of ZW sex chromosomes together indicate that the formation of ZW sex chromosomes was a major transition associated with early caenophidian evolution.

The rattlesnake W chromosome has evolved a unique complement of features since the establishment of recombination suppression. A particularly striking feature is relatively extreme GC-richness, which contrasts sharply with sex-limited chromosomes of other amniote species (Fig. 3). Chromosomes lacking sexual recombination are expected to have a lower frequency of GC nucleotides than recombining genomic regions because GC-biased gene conversion is expected to be less prevalent or absent altogether (Eyre-Walker and Hurst 2001; Galtier et al. 2001; Duret and Galtier 2009). Indeed, comparisons reveal a pattern of low GC content on the sex-limited chromosome across other ZW and XY amniote vertebrates (Fig. 3). We find that GC-richness on the rattlesnake W chromosome coincides with an abundance of GC-rich mdg4 elements, which comprise over 23% of the rattlesnake W chromosome (Figs. 3–4; Supplementary Table S5). The relative abundance of full-length mdg4 elements on the W chromosome further suggests that mdg4-driven GC-richness has evolved since the onset of recombination suppression between the sex chromosomes, which promoted the retention of these potentially transpositionally-active elements.

Our finding that mdg4 element insertion has played a dominant role in the evolution and divergence of the rattlesnake W chromosome raises questions about what features of this particular element may have favored its invasion of the W, and what the biological ramifications of this invasion may be. Studies in *Drosophila* have indicated that mdg4 elements may accumulate in specific genomic regions through a feed-forward process in which the presence of an existing element significantly increases the probability of insertion of additional new elements into adjacent genomic sequences (Labrador et al. 2008). Evidence that the rattlesnake W chromosome is a major hotspot for mdg4 insertion is consistent with this model, which may partially explain why mdg4 elements have accumulated at such high frequencies. Mdg4 elements are known to be transmitted as retrovirus-like pro-viral particles (Labrador et al. 2008), which suggests they are highly capable of horizontal transfer. Unlike most LTR- and retroviral-like elements that target euchromatic open chromatin (Scherdin et al. 1990; Schröder et al. 2002; Wu et al. 2003), mdg4 elements tend to be site-specific and target heterochromatic regions (Labrador et al. 2008). In *Drosophila*, mdg4 elements are also transmitted through female oocytes, and insert in germline tissue via a piRNA-

dependent mechanism, suggesting their potential to preferentially target the heterochromatic W chromosome and female germline tissues (Dej et al. 1998). These elements are also known to modify chromatin and to form insulator sequences (i.e., elements with enhancer-blocking activity; Labrador et al. 2008). It is thus plausible that mdg4 elements could have contributed to the suppression of recombination between regions of the Z and W chromosomes and associated divergence of Z versus W gametolog expression via insertion of insulator sequences on the W chromosome. Future comparative studies that integrate data on gene expression, chromatin accessibility, and chromatin structure (e.g., ATACseq and Hi-C) would be valuable for testing these hypotheses regarding the potential functional ramifications of mdg4 element insertions on the W chromosome.

In addition to patterns of GC-richness driven by mdg4 retroelements, intrachromosomal gene conversion also appears to have contributed to the GC-richness of the rattlesnake W chromosome and may represent a mechanism to counteract gene decay on the W chromosome. Using simulations, Marais et al. (2010) showed that gene conversion between Y-linked gene copies can be beneficial for opposing Muller's Ratchet, with greater gene copy number providing an enhanced benefit. Other empirical examples also show a prevalence of gene conversion between amplified gene families on Y and W chromosomes (e.g., Rozen et al. 2003; Backström et al. 2005; Connallon and Clark 2010; Rogers et al. 2021) – the taxonomically widespread nature of gene conversion on sex-limited chromosomes indeed suggests that it may be advantageous in tempering rates of gene decay. Several lines of evidence support that the counter-degeneration effects of gene conversion are relevant to snake W chromosome evolution. Most generally, the observation of W genes with higher GC3 than matched Z gametologs (Fig. 6; Supplementary Table S3) suggests the presence of GC-biased gene conversion, analogous to the evolution of the *HINTW* gene family on the avian W chromosome (Ceplitis and Ellegren 2004; Backström et al. 2005; Nam and Ellegren 2008). Further, we estimate that the recent stratum has experienced a slower rate of gene decay than older strata, combined with a greater number of W-specific gene duplications and higher d_S values on average (Fig. 6). Based on this complement of patterns, we hypothesize that gene duplications have provided an evolutionary substrate for intrachromosomal gene conversion across the W chromosome, especially within the recent stratum. It is also possible that the mdg4-dominated repeat landscape of the W chromosome has further enhanced the propensity for intrachromosomal gene conversion.

Retroelements and the evolution of sex-linked mutational load

Among the most distinctive features of the rattlesnake W chromosome is that it is more than 80% repetitive, far surpassing the repeat content of autosomes and the Z chromosome (Fig. 3; Supplementary Table S5; Schield et al. 2019). In contrast, Singchat et al. (2020) reported much lower repeat content

based on the previously published genome assembly for the Indian cobra (*Naja naja*) that contained a scaffold identified as the W chromosome (Super-Scaffold_1000010; Suryamohan et al. 2020). The surprising contrast in repeat landscapes between the two species, along with other preliminary comparative analyses, led us to question the W-linkage of Indian cobra Super-Scaffold_1000010, and we performed a series of analyses designed to test whether the scaffold is instead autosomal (and thus misidentified as the W chromosome). Indeed, comparative mapping of male and female reads combined with synteny analyses indicate that this Indian cobra scaffold represents a fragment of chromosome 6 (Supplementary Appendix; Supplementary Fig. S2A-B). Based on these results, we conclude that prior characterization of this presumed W chromosome in the Indian cobra was in error, and instead described characteristics of an autosome wrongly inferred to represent the W chromosome in that species (a list of candidate W-linked Indian cobra scaffolds identified in this study is provided in the Supplementary Data). As such, the repeat-poor inference reported in Singchat et al. (2020) is not representative of the snake W chromosome repeat landscape, and our findings presented here for the rattlesnake W chromosome provide, to our knowledge, the first accurate characterization of the repeat element accumulation and composition on a snake W chromosome.

Accumulation of repeat content on the rattlesnake W chromosome is consistent with the large body of empirical evidence that TE activity and accumulation drives the evolution of highly repetitive sex-limited chromosomes (Bachtrog 2013; Smeds et al. 2015; Tomaszkiewicz et al. 2017; Peona et al. 2021). Here, transposable element insertions likely persist at a higher frequency on the W chromosome because selection is less efficient against slightly deleterious mutations due to the magnified effects of genetic drift and selective interference – ultimately a consequence of recombination arrest (Charlesworth et al. 1986; Charlesworth and Charlesworth 2000). While mechanisms (e.g., gene conversion) may exist to oppose genetic decay, at a broad scale their effects appear to be limited relative to the dominant impacts of retroelements, including full length elements capable of transpositional activity, in shaping the mutational landscape of the W chromosome. Moreover, the relevance of these factors supports the view that the W chromosome is the product of antagonistic evolutionary mechanisms, culminating in a fine balance between decay and retention of ancestral genomic regions and genes.

The repetitive nature of sex-limited chromosomes may have important implications for sex-biased mutational load and genome function though epistatic effects on chromatin structure. The observation that W(Y) chromosomes act as reservoirs for transposable elements that have large-scale effects on gene expression during aging due to breakdowns in heterochromatin has led to the 'toxic Y' hypothesis (Brown et al. 2020b; Nguyen and Bachtrog 2020; Wei et al. 2020), and spurred the development of metrics to

quantify the toxicity of sex-limited chromosomes and the degree to which they act as transposable element refugia (Peona et al. 2021). We find that the rattlesnake W chromosome is highly enriched for retroelements, including multiple retroviral-like elements (Fig. 4; Table 2; Supplementary Table S6), which are similarly abundant on the W chromosome in bird species (Peona et al. 2021). These include a disproportionate number of full-length elements that have likely been protected from purifying selection due to selective interference on the W chromosome, providing the first evidence that the snake W chromosome acts as a refugium for self-replicating transposable elements. This suggests the W chromosome may be a source of higher mutational load in female snakes, with possible downstream impacts on gene regulation and genomic instability analogous to those reported for birds (Peona et al. 2021) and *Drosophila* (Nguyen and Bachtrog 2020; Wei et al. 2020). The potential for genetic incompatibilities to arise on the W chromosome may have additional downstream impacts relevant to speciation, as the build-up of mutational load could drive reproductive isolation between snake lineages, similar to that suggested in birds (Peona et al. 2021). Intriguingly, we do not find support for female-biased toxicity based on the toxicity index, as both Z and W chromosomes are enriched for full-length elements relative to autosomes, yielding an index near zero (see Results). An important caveat to consider, however, is that the fragmentary nature of our W chromosome assembly may bias our inference of full-length elements (i.e., it likely represents a lower-bound on the true number of full-length elements), compared to the contiguous assembly of the Z chromosome and autosomes (Table 1; Schield et al. 2019). It therefore remains an open question if there is truly no sex bias in 'toxicity', and additional analysis of the presence and activity of full-length elements based on contiguous W chromosome assemblies will be useful for further testing hypotheses related to female-specific mutational load.

Specialization of the W chromosome for female function

Despite substantial degeneration of the rattlesnake W chromosome, it has retained a subset of genes that may be important for female-specific function (Fig. 5). Indeed, our characterization of W-linked genes indicates retention of genes involved in reproduction, development, and transcriptional regulation. Specifically, the W chromosome is enriched for genes involved in embryonic development (i.e., skeletal and nervous system development; Fig. 5), and also houses multiple genes involved in early developmental differentiation (e.g., HOX genes), programmed cell death, and chromatin regulation. One particular retained gene on the W chromosome, *LMTK3*, is notable because it functions as a positive regulator of the estrogen receptor *ESR1* (Giamas et al. 2011), which is known to play a role in promoting female-biased gene expression in vertebrates (Nilsson et al. 2001; Rice et al. 2017). A previous study found evidence for an evolutionary increase in the abundance of estrogen response elements (cis-regulatory elements bound by ESR1) on the rattlesnake Z chromosome that may facilitate gene- or

region-specific dosage compensation (Schield et al. 2019). It is therefore plausible that survival of the *LMTK3* W gametolog was favored by selection because it is necessary for activation of *ESR1* related to its role in estrogen-driven upregulation of dose-sensitive genes in females.

Several genes involved in immune function have also translocated to the W chromosome and or experienced W-specific amplification through apparent gene duplication (e.g., *H2Q9*, *HA1F*, *RBM14*, and *RT1B*). Why translocation and pronounced duplication of such genes has evolved is unclear, but it is logical that an associated increase in immune function is beneficial to female survival and reproduction. In the case of W-specific duplications of immune genes, one possibility is that selection has favored amplification to compensate for lower expression on the highly heterochromatic W chromosome, similar to explanations for expanded gene families on the *Drosophila miranda* neo-Y chromosome (Bachtrog 2006, 2020). An alternative explanation, analogous to the hypothesized role of *HINTW* expansion in birds (Backström et al. 2005), is that duplication is simply favorable as a mechanism to safeguard against genetic decay and ultimate loss of immune genes on the W chromosome.

Conclusion

Snakes inspired foundational work that established prevailing views of sex chromosome evolution (Ohno 1967), yet characterizations of snake sex chromosome genomic structure and function have since lagged behind those of other vertebrate groups, leaving a major gap in our understanding of the forces that shape the evolution of sex chromosomes in vertebrates. Here, we provide the first detailed analysis of the W chromosome from a caenophidian snake, the prairie rattlesnake. Our findings reveal a unique complement of structural features on the W chromosome that bear similarities and contrasts with other species, providing an expanded view of vertebrate sex chromosome evolution. The diverse and often conflicting dynamics between specialization of the W chromosome for vital female-specific functions, retention of critical genes, retroelement proliferation, and ongoing genetic decay highlight the complexity of unique evolutionary forces that shape sex chromosomes. These findings also raise new questions for future comparative study, including: 1) how has recombination suppression proceeded in other caenophidian lineages, 2) has specialization for female function followed independent evolutionary trajectories in distinct snake ZW and XY systems, 3) how have mdg4 elements contributed to recombination suppression, chromatin structure, and divergence of gene expression and function on the W chromosome, and 4) how does female-specific mutational load and the potential 'toxicity' of the W chromosome impact genomic regulation and are these effects relevant to speciation? Finally, though linked-read sequencing enabled the assembly of W chromosome scaffolds, these remain fragmentary likely due to high heterochromatin and repetitive content of the chromosome. Additional scaffolding using long-read

technologies and Hi-C chromatin contact data (especially for trios of parents and offspring) holds promise for increasing the contiguity of highly-degenerated sex chromosome assemblies in the rattlesnake and other snake species.

Materials and Methods

Assembly of Prairie Rattlesnake W chromosome scaffolds

We used the prairie rattlesnake reference genome assembly and annotation (Schield et al. 2019a; NCBI BioProject PRJNA413201) for analyses of autosomes and the Z chromosome. To generate sequence data for the W chromosome, we sampled a female prairie rattlesnake from the same population as the male genome animal. Liver tissue was snap-frozen in liquid nitrogen and stored at -80° C. Genomic DNA was extracted and a 10x Genomics Chromium library was prepared to enable linked-read sequencing on an Illumina NovaSeq 6000 using 150 bp paired-end reads. We assessed sequence quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and summarized the results using MultiQC (Ewels et al. 2016), which are shown in Supplementary Fig. S1. We performed genome sub-assembly using the 10x Genomics Supernova assembler v2.1.1 (Weisenfeld et al. 2017) specifying 560 million input reads based on calculations of recommended fold genome-coverage and the estimated genome size. We used the 'pseudohap2' option to generate scaffold sequences in order to retain as many potentially W-linked scaffolds as possible.

We identified W-linked scaffolds by first performing a homology search of all female scaffolds against the male genome assembly using MashMap (Jain et al. 2018), using a 95% identity threshold and the one-to-one filtering option. We removed high-similarity hits to autosomes and scaffolds that were less than 5 kb in length and retained hits to the Z chromosome and also any scaffolds that did not have significant sequence similarity to the male reference (i.e., female-specific scaffolds). This procedure is expected to enrich for a list of W-linked, Z-linked, and pseudoautosomal region (PAR) scaffolds from the female genome assembly. We then filtered scaffolds with hits to the 6.8 Mb PAR region, leaving 15,254 putatively sex-linked scaffolds.

We then used comparative mapping of whole genome resequencing reads from both sexes to distinguish W- and Z-linked female scaffolds. We quality trimmed reads (Supplementary Table S1) using Trimmomatic v0.39 (Bolger et al. 2014) with the settings LEADING:20 TRAILING:20 MINLEN:32 AVGQUAL:30. We mapped filtered reads to all female scaffolds using default settings in BWA mem

v0.7.17 (Li and Durbin 2009), then removed mapped reads with quality scores below Q30 using Samtools v1.1 (Li et al. 2009). We calculated sequencing depth per scaffold using Mosdepth (Pedersen and Quinlan 2018) with default parameters. We then used the distribution of read depths from scaffolds with best hits to autosomes to calculate median autosomal depth per sex.

Statistical methods used to identify W-linked sequence based on comparative mapping have different associated power and false-positive rates, with the potential to spuriously assign sex-linkage to autosomal regions or vice versa. To address this concern, we performed comparative mapping analyses in chicken (*Gallus gallus*) using available female and male resequencing data (Cortez et al. 2014; NCBI SRA accessions SRR958465 and SRR958466) mapped to version 6 of the *Gallus gallus* genome, which includes an assembled W chromosome (Hillier et al. 2004; Bellott et al. 2017). We filtered and mapped reads following the methods described for rattlesnake data and calculated mean read depth in genomic windows using Mosdepth. The mean length of female rattlesnake scaffolds was 10,013.8 bp, so we measured read depth in the chicken using 10 kb windows for comparison. We then counted the number of true and false positives when identifying putative W-linked chicken regions using two alternative approaches. First, we calculated

$$log_2\left(\frac{female\ autosome - normalized\ depth}{male\ autosome - normalized\ depth}\right)$$

per genomic window, where autosome-normalized depth is equal to the mean read depth per window divided by the autosomal median read depth for each sex. We refer to this measure as 'log_2FM'. The expectation for autosomes is $Log_2FM = 0$. Z-linked sequences, on the other hand, are expected to have $Log_2FM = -1$ and W-linked sequences are expected to have $Log_2FM \sim 2$. To accommodate variance in read depths among genomic windows, we identified putative W-linked windows using a threshold of $log_2FM \geq 1$. Second, we calculated the autosome-normalized proportion of female read depth per genomic window. Each window with a female read depth proportion greater than the third quartile threshold plus 1.5 times the inter-quartile range (i.e., Q3 + 1.5*IQR) was considered putatively W-linked. Each approach had high power to detect W-linked genomic windows (log_2FM power = 0.954; Q3 + 1.5*IQR power = 0.97). Despite having slightly greater power, the Q3 + 1.5*IQR method also suffered nearly an order of magnitude higher false-positive rate (log_2FM false-positive rate = 0.0049; Q3 + 1.5*IQR false-positive rate = 0.025). Based on the mapping experiment in chicken, we used the more conservative log_2FM threshold to identify W-linked scaffolds in the prairie rattlesnake. We calculated log_2FM per female scaffold, identified putative W-linked sequences using the $Log_2FM \geq 1$ threshold, then cross-

referenced these with our list of putative sex-linked scaffolds from the initial homology procedure using the male reference genome. This yielded 2,139 female scaffolds with overlapping evidence of W-linkage (Table 1).

Annotation of the W chromosome

We annotated repeat elements and protein coding genes to facilitate comparative analyses of sex-linked genes and to characterize the composition and structure of the W chromosome. Repeat elements were annotated using RepeatMasker v4.0.8 (Smit et al. 2015), leveraging a Bov-B/CR1 LINE homology database generated using multiple squamate genomes (Pasquesi et al. 2018), tetrapod elements included in RepBase release 20181026 (Bao et al. 2015), and known and unknown consensus elements from a snake-specific repeat library (Castoe et al. 2013; Pasquesi et al. 2018; Schield et al. 2019).

We used MAKER v2.31.10 (Cantarel et al. 2008) to produce an initial annotation of protein-coding genes in W-linked scaffolds using empirical evidence for gene prediction (settings est2genome=1 and protein2genome=1). Empirical evidence included the *de novo* transcriptome assembly used in the male prairie rattlesnake genome assembly (Schield et al. 2019) combined with a *de novo* transcript assembly derived from 18 tissues from female prairie rattlesnakes (Supplementary Table S2). Tissue samples were collected and immediately snap frozen in liquid nitrogen and stored at -80° C. Total RNA was extracted using Trizol and poly-A selected mRNA libraries were generated and sequenced on an Illumina HiSeq 2500 using 100 bp paired-end reads. Reads were randomly subsampled using the sample tool in seqtk v1.3-r106 (www.github.com/lh3/seqtk) to retain two million reads per sample. We then generated the transcriptome assembly using Trinity release 2014-07-17 (Grabherr et al. 2011) with the --trimmomatic flag to incorporate upfront quality trimming by Trimmomatic (Bolger et al. 2014) with default settings.

The *de novo* transcriptomes were supplied to the argument 'est', along with protein datasets for all annotated protein-coding genes of *Anolis carolinensis* (Alfoldi et al. 2011), *Python molurus bivittatus* (Castoe et al. 2013), *Thamnophis sirtalis* (Perry et al. 2018), *Ophiophagus Hannah* (Vonk et al. 2013), and *Deinagkistrodon acutus* (Yin et al. 2016) supplied to the argument 'protein'. An initial round of MAKER was run with default settings, except that we specified max_dna_len=300000 and split_hit=20000. The resulting 386 gene models were used to optimize gene prediction parameters in Augustus v3.2.3 (Stanke and Morgenstern 2005), which we used for gene prediction in a second run of MAKER that was identical to the first, except for specifying est2genome=0 and protein2genome=0 and with Augustus parameters supplied to the 'augustus_species' setting. The resulting annotation contained 213 protein-coding gene models (Table 1).

We further used transcript data to improve the contiguity of the W chromosome assembly. We gathered the existing assembly together with gene models from MAKER in GFF3 format and mapped all female RNAseq read data used in the *de novo* transcriptome using default settings in BWA mem (Li and Durbin 2009). These data were then analyzed using AGOUTI v. 0.3.3-dirty (Zhang et al. 2016) with default settings except for minMQ=20 and maxFracMM=0.05. AGOUTI further scaffolded 206 of the 2,139 W chromosome scaffolds into 94 final scaffolds, resulting in a final assembly of 2,027 scaffolds with an N50 = 13,252 bp. We again performed repeat and gene annotations following the same procedures as above, resulting in 219 protein models after the second run of MAKER (Table 1). We ascribed gene IDs based on homology for 134 of these gene models using a reciprocal best BLAST (Altschul et al. 1990; with an e-value threshold of 1×10^{-5}) and stringent one-way BLAST (with an e-value threshold of 1×10^{-8}) searches against protein sequences from NCBI for *Anolis*, *Python*, and *Thamnophis*.

Sex chromosome homology and identification of evolutionary strata

Previous investigations have hypothesized the presence of at least two evolutionary strata between Z and W chromosomes in caenophidian snakes (Vicoso et al. 2013; Yin et al. 2016; Schield et al. 2019). Comparisons of Z and W-linked gametologs in the context of the Z chromosome assembly for prairie rattlesnake provide the opportunity to examine the structure, age, and number of evolutionary strata in greater detail. To establish the structure of the assembled W chromosome in relation to the Z chromosome, we used MashMap (Jain et al. 2018) to anchor W-linked scaffolds with sufficient sequence similarity to homologous regions of the Z chromosome. We specified '-f one-to-one' to limit matches to reciprocal best hits, specified a minimum search length of 5 kb, and filtered hits with lower than 90% sequence similarity.

We identified ZW gametologs using tBLASTx (Altschul et al. 1990) searches between annotated coding sequences (CDS) with the setting '-max_hsps 1' and an e-value threshold of 1 × 10-5. We then extracted reciprocal best BLAST hits as 1:1 gametologs, yielding 125 1:1 gametolog pairs. We translated nucleotide sequences for each pair and aligned amino acid sequences using Clustal Omega (Sievers et al. 2011), then used PAL2NAL (Suyama et al. 2006) to generate codon-based nucleotide alignments. We estimated synonymous (d_S) and nonsynonymous (d_N) divergence between gametologs using the codon model in PAML (Yang 2007), and filtered pairs with d_S above 2 and below 0.001. To transform synonymous divergence estimates to divergence times, we calculated a lineage-specific mutation rate based on divergence between *Crotalus* and *Anolis* autosomal genes. We identified orthologous genes using a reciprocal best BLAST search, which yielded 11,277 1:1 autosomal orthologs. We aligned orthologs and calculated d_S as described above, then obtained a mutation rate estimate by dividing median

d_S (0.94) by 167 million years, the estimated divergence time for *Crotalus* and *Anolis* from TimeTree (Kumar et al. 2017), yielding a lineage-specific autosomal mutation rate of 2.8×10^{-9} per site per year. To account for sex-linked mutational biases, we multiplied this rate by the Z chromosome:autosome mutation rate ratio (μ_Z/μ_A) of 1.1 estimated in Schield et al. (2021). The W chromosome is present only in females and therefore has a female-specific mutation rate. While this rate is unknown in rattlesnakes, estimates of the male to female mutation rate ratio in birds, which have similar estimates of male-biased mutation rates to snakes (Vicoso et al. 2013; Schield et al. 2021), range from two to four (Nam and Ellegren 2008). To accommodate female-specific mutation on the W chromosome, we conservatively used a female mutation rate equal to one-fourth of the autosomal rate, adding this to the inferred rate for the Z chromosome to produce a combined sex-linked divergence rate ($3.1 \times 10^{-9} + 0.7 \times 10^{-9} = 3.8 \times 10^{-9}$). We transformed d_S estimates for ZW gametologs to time using this rate, then calculated the mean divergence time per stratum. For comparison, and to obtain a range of approximate divergence times per stratum, we repeated this procedure using a squamate mutation rate derived from four-fold degenerate sites (2.4×10^{-9}; Green et al. 2014), which, after conversion to the sex-linked divergence rate was equal to 3.24×10^{-9}.

We investigated whether the recent stratum is present in non-pitviper colubroids by mapping female and male western terrestrial garter snake (*Thamnophis elegans*, a colubrid) read data (Vicoso et al. 2013; Supplementary Table S1) to the prairie rattlesnake genome using bwa mem v0.1.17 (Li and Durbin 2009) after quality filtering reads using Trimmomatic v0.39 (Bolger et al. 2014) with settings described for other analyses above. We removed mapped reads with quality scores lower than Q30 using Samtools v1.1 (Li et al. 2009) and quantified read depths in 10 kb sliding windows using Mosdepth (Pedersen and Quinlan 2018). We calculated $\log_2$FM per window on the Z chromosome based on median autosomal read depths per sex, following our methods used for discovery of sex-linked genomic regions. We repeated these steps using resequencing data from three pitviper species: five-pace viper (*Deinagkistrodon acutus*; Yin et al. 2016), pygmy rattlesnake (*Sistrurus miliarius*; Vicoso et al. 2013), and prairie rattlesnake (Supplementary Table S1). If recombination suppression in the recent stratum predates the common ancestor of the garter snake and pitvipers, $\log_2$FM would be intermediate between -1 and 0 in this region, corresponding to expectations for Z-linked and autosomal regions, respectively. Alternatively, $\log_2$FM values consistently near -1 would indicate that garter snake sex chromosomes experienced recombination suppression in this region independently of pitvipers.

GC and repeat content on the W chromosome

We measured GC content, CpG content, and repeat content on the W chromosome using custom Python scripts (https://github.com/drewschield/rattlesnake_w_chromosome). GC content was measured as the

proportion of G and C nucleotides per scaffold after removal of ambiguous bases. Similarly, CpG content was measured as the proportion of CG dinucleotides, and repeat content was measured as the proportion of bases per scaffold annotated as repeats. We also measured GC and CpG content on the prairie rattlesnake Z chromosome and autosomes and across the genomes of chicken (*Gallus gallus* version GRCc6a; Hillier et al. 2004; Bellott et al. 2017), zebra finch (*Taeniopygia guttata* version bTaeGut2.pat.W.v2; Warren et al. 2010), human (*Homo sapiens* version GRCh38.p13; Venter et al. 2001) and house mouse (*Mus musculus* version GRCm39; Waterston et al. 2002) in 10 kb windows. The reference genomes for each of these species include representative assembled scaffolds for Z(X) and W(Y) sex chromosomes.

Measures of repeat content were based on the repeat annotation described above, and we used the CpG-corrected Kimura 2-parameter distance as a measure of the age distribution for specific element families. Several TE families were abundant on the W chromosome, including multiple LTR retrotransposons – mdg4, L1 LINEs, CR1/L3 LINEs, and ERVs (see Results). In addition to total repeat content, we measured the proportion of bases from these TE families per W-linked scaffold, and in 10 kb windows on the Z chromosome and autosomes. We compared GC and CpG content of mdg4, L1, and CR1/L3 elements to their densities on the autosomes, Z chromosome, and W chromosome, and examined the distributions of GC and CpG content for each of these families on the W chromosome.

Analysis of retroelements

We compared the frequencies of repeat elements on autosomes and the sex chromosomes to test the hypothesis that there is a uniform distribution of repeats across the genome. Following Peona et al. (2021), we first determined the expected repeat-derived bp by assuming that the total length of repeats is proportional to the total length of autosomes, Z, and W chromosomes after removing ambiguous bases. We then compared observed values to the expected values to calculate the refugium index for each chromosome class using the formula:

$$Refugium\ Index = \frac{\%TE_{obs} - \%TE_{exp}}{\%TE_{exp}}.$$

The refugium index provides information on whether the autosomes and sex chromosomes have a depletion (refugium index < 0) or excess (refugium index > 0) of repeat elements compared to a uniform distribution. We tested for significant deviation from a uniform distribution using χ^2 tests. We repeated

these analyses for specific abundant retroelements on the W chromosome (i.e., mdg4, ERVs, L1 LINEs, CR1/L3 elements).

To compare frequencies of long terminal repeat (LTR) retroelements capable of self-replication across the genome, we identified full-length elements with intact LTR and protein domains (fl-LTRs) using LTRharvest (Ellinghaus et al. 2008) and LTRdigest (Steinbiss et al. 2009). We filtered results using Pfam (Mistry et al. 2021) and GyDB (Llorens et al. 2010) hidden Markov model profiles for LTR retrotransposon proteins. We then compared observed and expected numbers of fl-LTRs to calculate the refugium index for autosomes and the sex chromosomes.

To examine evidence of sex differences in toxicity on the basis of homogametic and heterogametic linkage of fl-LTRs, we calculated the toxicity index as defined in Peona et al. (2021):

$$Toxicity\ index = \frac{2n_{het} - 2n_{hom}}{2n_{hom}}.$$

Here, $2n_{\text{het}}$ is equal to the number of diploid fl-LTRs in the heterogametic sex (i.e., 2 × autosomes + 1 × Z + 1 × W) and $2n_{\text{hom}}$ equals the number in the homogametic sex (2 × autosomes + 2 × Z). An index equal to 0 indicates no sex bias in toxicity, whereas a toxicity index > 0 indicates female-biased toxicity and < 0 indicates male-biased toxicity in the case of ZW snakes.

Gene expression and characterization of W-linked genes

We quantified gene expression of W-linked genes, with the expectation that genes that have remained functional since recombination suppression are expressed in female tissues. We mapped mRNA-seq datasets for ovary and testes (NCBI SRA accession PRJNA477004) and for two male and female liver and kidney samples (10 samples total; Supplementary Table S2) to the updated genome annotation using STAR v2.7.1a (Dobin et al. 2013) with the flags --outFilterMultimapNmax 1, --outFilterMismatchNmax 1, --outFilterMismatchNoverLmax 0.1, and --twoPassMode Basic, then quantified raw read counts using featureCounts v2.0.1 (Liao et al. 2013). Raw counts from featureCounts were used to assess whether genes on the W chromosome had "detectable" expression in female and male samples, defined as a raw count > 0 in at least one sample of a given sex. Genes with detected expression in female samples and no detected expression in male samples were considered to have female-specific expression. To further assess female-specific expression, we performed pairwise comparisons between female and male samples using DEseq2 v1.30.1 (Love et al. 2014) and conducted *p*-value correction using IHW v1.18.0 (Ignatiadis et al. 2016) with baseMean expression as the covariate. Female-specific genes were defined as those with

significant differential expression (IHW p-value < 0.05) and higher expression in female samples compared to males (i.e., $\log_2$FoldChange > 0). TPM-normalized counts were calculated in R and averaged across males and females to provide representative expression values for each sex. Heatmaps were generated using pheatmap v1.0.12 (Kolde 2012) in R.

We functionally annotated genes on the W chromosome using gene ontology (GO) term analysis in WebGestaltR (Liao et al. 2019; https://github.com/bzhanglab/WebGestaltR), with W-linked genes as the foreground and annotated genes on other chromosomes as the background gene set. Using the overrepresentation analysis method, we selected the top 20 overrepresented terms from combined non-redundant biological process, cellular component, and molecular function GO-term databases, and a combined KEGG (Kanehisa and Goto 2000), Panther (Mi and Thomas 2009; Mi et al. 2021), Reactome (Jassal et al. 2020), and Wikipathway (Martens et al. 2021) pathway database. We tested for enrichment of W-linked genes for specific terms using Fisher's exact tests and defined significantly overrepresented terms as those with p-values below 0.05 after FDR correction for multiple tests. We characterized protein classes for W-linked genes using PantherDB (Mi et al. 2019). This method was also used to characterize W-specific duplicated and translocated genes described below.

Analysis of W-linked genes: survival, duplication, selection, and gene conversion

Previous evidence suggests disproportionate survival of W-linked genes in the recent stratum (Schield et al. 2019). To test this hypothesis, we calculated a decay rate for stratum 1 (i.e., older strata) by subtracting the number of W genes from the number of Z genes present in stratum 1, divided by the inferred divergence time for the stratum. We compared this rate to one for the recent stratum to test whether genes have decayed at similar rates in the time since recombination suppression in both strata. To test for evidence of autosome-to-W chromosome translocations, we used a reciprocal best BLAST between W chromosome and *Anolis* CDS sequences. *Anolis* chromosome 6 is homologous to the colubroid ZW chromosomes, therefore we retained any hits to other chromosomes that also did not have a matched Z chromosome gametolog as candidate autosomal translocations.

We tested for W-specific gene duplications by mapping female resequencing data (n = 4; Supplementary Table S1; Schield et al. 2021, 2022) to autosomes, the Z chromosome, and the W chromosome using BWA mem v0.7.17 (Li and Durbin 2009), after quality filtering reads using Trimmomatic v0.39 (Bolger et al. 2014) with settings described above. We calculated mean read depth per gene (within exon boundaries) using Mosdepth (Pedersen and Quinlan 2018) and determined the median among autosomal genes per female. We then quantified relative W-linked gene read depth by dividing mean depth per gene

by the autosomal median in each female after accounting for haploidy of the W chromosome. This provided relative depths per gene scaled to a value of 1. We inferred copy-number as the mean of relative depth values observed across females per W-linked gene.

We compared rates of sequence divergence between prairie rattlesnake and five-pace viper Z and W chromosome orthologs, respectively, following the procedure described above for prairie rattlesnake ZW gametologs. Briefly, we identified sex-linked orthologs between the two species, then aligned orthologs using Clustal Omega (Sievers et al. 2011) and PAL2NAL (Suyama et al. 2006). We used the codon model in PAML (Yang 2007) to estimate d_N, d_S, and d_N/d_S separately for the Z chromosome and W chromosome. We compared rates of nonsynonymous and synonymous evolution between the sex chromosomes to detect whether expected differences in the efficiency of natural selection are present on the female-specific W chromosome. Cases of gene conversion within genes on sex-limited chromosomes have been documented (e.g., Backström et al. 2005), along with an associated increase in GC bases due to GC-biased gene conversion. To test for evidence of gene conversion within W-linked genes (i.e., intrachromosomal recombination), we compared the proportion of GC bases at third codon positions (GC3) between matched Z and W gametologs in the prairie rattlesnake.

Statistical analysis

We performed statistical analyses comparing distributions of genomic variation and divergence in R (R Core Team 2017). Prior to performing pairwise comparisons between variables, we tested for normality of d_S, d_N/d_S, GC content, CpG content, repeat content (including specific repeat element families), and GC3 distributions using Shapiro-Wilk tests. We further evaluated homogeneity of variances between distributions using Levene's tests. All distributions rejected normality (p-values < 0.05) with the exception of GC3. Accordingly, we tested for significant differences in pairwise comparisons of d_S, d_N/d_S, GC content, CpG content, and repeat content distributions on the sex chromosomes and autosomes using Mann-Whitney U tests and tested for associations between variables using Spearman's rank order correlation coefficients. We calculated partial correlation coefficients between GC content, CpG content, and repeat content while controlling for correlation among sets of variables using the 'ppcor' R package (Kim 2015). Because GC3 distributions were not significantly different from a normal distribution, we performed pairwise comparisons using Welch's two-sample t-tests. All distributions supported the hypothesis of equal variances (i.e., Levene's test p-values > 0.05) with the exception of d_S estimates between ZW gametologs among inferred evolutionary strata (see Results).

Acknowledgments

We thank Andrew Hillhouse at Texas A&M University for assistance with library preparation and sequencing. We thank Rebecca Safran, Zach Laubach, and Heather Kenny for helpful discussions about this work. This work was supported by the National Science Foundation grant numbers DBI-1906188 to DRS, OPP-2138649 to BWP, DEB-1812310 to DCC, and DEB-1655571 to TAC and the University of Northern Colorado Research Dissemination and Faculty Development grant to SPM. All procedures using animals or animal tissue were performed according to the University of Colorado Institutional Animal Care and Use Committee (IACUC) protocols 0901C-SM-MLChick-12 and 1302D-SM-S-16.

Data Availability

The W chromosome assembly generated in this study is available from NCBI Genbank under accession PRJNA853338. RNAseq data generated in this study are available from the NCBI short-read archive under accession PRJNA477004. Gene and repeat annotations are available at https://github.com/drewschield/rattlesnake_w_chromosome/resources/annotation. The computational workflow and analysis scripts are available at https://github.com/drewschield/rattlesnake_w_chromosome.

References

Abbott, J. K., A. K. Nordén, and B. Hansson. 2017. Sex chromosome evolution: historical insights and future perspectives. *Proc R Soc B* 284:20162806.

Alfoldi, J., F. Di Palma, M. Grabherr, C. Williams, L. Kong, E. Mauceli, P. Russell, C. B. Lowe, R. E. Glor, J. D. Jaffe, D. A. Ray, S. Boissinot, A. M. Shedlock, C. Botka, T. A. Castoe, J. K. Colbourne, M. K. Fujita, R. G. Moreno, B. F. ten Hallers, D. Haussler, A. Heger, D. Heiman, D. E. Janes, J. Johnson, P. J. de Jong, M. Y. Koriabine, M. Lara, P. A. Novick, C. L. Organ, S. E. Peach, S. Poe, D. D. Pollock, K. de Queiroz, T. Sanger, S. Searle, J. D. Smith, Z. Smith, R. Swofford, J. Turner-Maier, J. Wade, S. Young, A. Zadissa, S. V Edwards, T. C. Glenn, C. J. Schneider, J. B. Losos, E. S. Lander, M. Breen, C. P. Ponting, and K. Lindblad-Toh. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477:587–591.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.

Augstenová, B., M. Johnson Pokorná, M. Altmanová, D. Frynta, M. Rovatsos, and L. Kratochvíl. 2018. ZW, XY, and yet ZW: Sex chromosome evolution in snakes even more complicated. *Evolution* 72:1701–1707.

Axelsson, E., N. G. C. Smith, H. Sundström, S. Berlin, and H. Ellegren. 2004. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Mol Biol Evol* 21:1538–1547.

Bachtrog, D. 2003. Accumulation of Spock and Worf, two novel non-LTR retrotransposons, on the neo-Y chromosome of *Drosophila miranda*. *Mol Biol Evol* 20:173–181.

Bachtrog, D. 2006. Expression profile of a degenerating neo-Y chromosome in *Drosophila*. *Curr Biol* 16:1694–1699.

Bachtrog, D. 2005. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. *Genome Res* 15:1393–1401.

Bachtrog, D. 2020. The Y Chromosome as a Battleground for Intragenomic Conflict. *Trends Genet* 36:510–522.

Bachtrog, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* 14:113–124.

Bachtrog, D., S. Mahajan, and R. Bracewell. 2019. Massive gene amplification on a recently formed *Drosophila* Y chromosome. *Nat Ecol Evol* 3:1587–1597.

Backström, N., H. Ceplitis, S. Berlin, and H. Ellegren. 2005. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. *Mol Biol Evol* 22:1992–1999.

Baker, R. J., J. J. Bull, and G. A. Mengden. 1972. Karyotypic Studies of 38 Species of North-American Snakes. *Copeia* 1972:257–265.

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

Becak, W., M. L. Beçak, H. R. S. Nazareth, and S. Ohno. 1964. Close karyological kinship between the reptilian suborder Serpentes and the class Aves. *Chromosoma* 15:606–617.

Bellott, D. W., H. Skaletsky, T.-J. Cho, L. Brown, D. Locke, N. Chen, S. Galkina, T. Pyntikova, N. Koutseva, and T. Graves. 2017. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat Genet* 49:387–394.

Bergero, R., and D. Charlesworth. 2009. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol* 24:94–102.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Brown, E. J., A. H. Nguyen, and D. Bachtrog. 2020a. The Drosophila Y chromosome affects heterochromatin integrity genome-wide. *Mol Biol Evol* 37:2808–2824.

Brown, E. J., A. H. Nguyen, and D. Bachtrog. 2020b. The Y chromosome may contribute to sex-specific ageing in Drosophila. *Nat Ecol Evol* 4:853–862.

Cantarel, B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado, and M. Yandell. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18:188–196.

Castoe, T. A., A. P. J. de Koning, K. T. Hall, D. C. Card, D. R. Schield, M. K. Fujita, R. P. Ruggiero, J. F. Degner, J. M. Daza, W. J. Gu, J. Reyes-Velasco, K. J. Shaney, J. M. Castoe, S. E. Fox, A. W. Poole, D. Polanco, J. Dobry, M. W. Vandewege, Q. Li, R. K. Schott, A. Kapusta, P. Minx, C. Feschotte, P. Uetz, D. A. Ray, F. G. Hoffmann, R. Bogden, E. N. Smith, B. S. W. Chang, F. J. Vonk, N. R. Casewell, C. V Henkel, M. K. Richardson, S. P. Mackessy, A. M. Bronikowsi, M. Yandell, W. C. Warren, S. M. Secor, and D. D. Pollock. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *PNAS* 110:20645–20650.

Ceplitis, H., and H. Ellegren. 2004. Adaptive molecular evolution of HINTW, a female-specific gene in birds. *Mol Biol Evol* 21:249–254.

Chang, C.-H., and A. M. Larracuente. 2019. Heterochromatin-enriched assemblies reveal the sequence and organization of the *Drosophila melanogaster* Y chromosome. *Genetics* 211:333–348.

Charlesworth, B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol* 6:149–162.

Charlesworth, B., and D. Charlesworth. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc B* 355:1563–1572.

Charlesworth, B., C. H. Langley, and W. Stephan. 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* 112:947–962.

Charlesworth, D., B. Charlesworth, and G. Marais. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128.

Cohen, M., M. Powers, C. O'Connell, and N. Kato. 1985. The nucleotide sequence of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. *Virology* 147:449–458.

Connallon, T., and A. G. Clark. 2010. Gene duplication, gene conversion and the evolution of the Y

chromosome. *Genetics* 186:277–286.

Cortez, D., R. Marin, D. Toledo-Flores, L. Froidevaux, A. Liechti, P. D. Waters, F. Gruetzner, and H. Kaessmann. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508:488–493.

Dej, K. J., T. Gerasimova, V. G. Corces, and J. D. Boeke. 1998. A hotspot for the *Drosophila* gypsy retroelement in the *ovo* locus. *Nucleic Acids Res* 26:4019–4024.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Duret, L., and N. Galtier. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* 10:285–311.

Ellinghaus, D., S. Kurtz, and U. Willhoeft. 2008. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* 9:1–14.

Ellison, C., and D. Bachtrog. 2019. Recurrent gene co-amplification on *Drosophila* X and Y chromosomes. *PLoS Genet* 15:e1008251.

Evans, L. H., M. Lavignon, M. Taylor, and A. S. M. Alamgir. 2003. Antigenic subclasses of polytropic murine leukemia virus (MLV) isolates reflect three distinct groups of endogenous polytropic MLV-related sequences in NFS/N mice. *J Virol* 77:10327–10338.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.

Eyre-Walker, A., and L. D. Hurst. 2001. The evolution of isochores. *Nat Rev Genet* 2:549–555.

Galtier, N., G. Piganeau, D. Mouchiroud, and L. Duret. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159:907–911.

Gamble, T., T. A. Castoe, S. V Nielsen, J. L. Banks, D. C. Card, D. R. Schield, G. W. Schuett, and W. Booth. 2017. The discovery of XY sex chromosomes in a boa and python. *Curr Biol* 27:2148–2153 e4.

Giamas, G., A. Filipović, J. Jacob, W. Messier, H. Zhang, D. Yang, W. Zhang, B. A. Shifa, A. Photiou, and C. Tralau-Stewart. 2011. Kinome screening for regulators of the estrogen receptor identifies LMTK3 as a new therapeutic target in breast cancer. *Nat Med* 17:715–719.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, and Q. Zeng. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* 29:644.

Graves, J. A. M. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124:901–914.

Green, R. E., E. L. Braun, J. Armstrong, D. Earl, N. Nguyen, G. Hickey, M. W. Vandewege, J. A. S. John, S. Capella-Gutiérrez, T. A. Castoe, C. Kern, M. K. Fujita, J. C. Opazo, J. Jurka, K. K. Kojima, J. Caballero, R. M. Hubley, A. F. Smit, R. N. Platt, C. A. Lavoie, M. P. Ramakodi, J. W. Finger Jr., A. Suh, S. R. Isberg, L. Miles, A. Y. Chong, W. Jaratlerdsiri, J. Gongora, C. Moran, A. Iriarte, J. McCormack, S. C. Burgess, S. V. Edwards, E. Lyons, C. Williams, M. Breen, J. T. Howard, C. R. Gresham, D. G. Peterson, J. Schmitz, D. D. Pollock, D. Haussler, E. W. Triplett, G. Zhang, N. Irie, E. D. Jarvis, C. A. Brochu, C. J. Schmidt, F. M. McCarthy, B. C. Faircloth, F. G. Hoffman, T. C.

Glenn, T. Gabaldón, B. Paten, and D. A. Ray. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346:1254449.

Handley, L.-J. L., H. Ceplitis, and H. Ellegren. 2004. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* 167:367–376.

Hillier, L. W., W. Miller, E. Birney, W. Warren, R. C. Hardison, C. P. Ponting, P. Bork, D. W. Burt, M. A. M. Groenen, and M. E. Delany. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.

Ignatiadis, N., B. Klaus, J. B. Zaugg, and W. Huber. 2016. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* 13:577–580.

Jain, C., S. Koren, A. Dilthey, A. M. Phillippy, and S. Aluru. 2018. A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics* 34:i748–i756.

Jassal, B., L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, and R. Haw. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res* 48:D498–D503.

Jiang, P.-P., D. L. Hartl, and B. Lemos. 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.

Kanehisa, M., and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30.

Kim, S. 2015. ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun Stat Appl Methods* 22:665.

Kolde, R. 2012. Pheatmap: pretty heatmaps. R Package. version 1.747.

Kruger, A. N., M. A. Brogley, J. L. Huizinga, J. M. Kidd, D. G. de Rooij, Y.-C. Hu, and J. L. Mueller. 2019. A neofunctionalized X-linked ampliconic gene family is essential for male fertility and equal sex ratio in mice. *Curr Biol* 29:3699–3706.

Kumar, S., G. Stecher, M. Suleski, and S. B. Hedges. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* 34:1812–1819.

Labrador, M., K. Sha, A. Li, and V. G. Corces. 2008. Insulator and Ovo proteins determine the frequency and specificity of insertion of the gypsy retrotransposon in *Drosophila melanogaster*. *Genetics* 180:1367–1378.

Lahn, B. T., and D. C. Page. 1999. Four evolutionary strata on the human X chromosome. *Science* 286:964–967.

Larson, E. L., E. E. K. Kopania, and J. M. Good. 2018. Spermatogenesis and the evolution of mammalian sex chromosomes. *Trends Genet* 34:722–732.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The sequence alignment/map format and SAMtools.

Bioinformatics 25:2078–2079.

Liao, Y., G. K. Smyth, and W. Shi. 2013. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.

Liao, Y., J. Wang, E. J. Jaehnig, Z. Shi, and B. Zhang. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* 47:199–205.

Llorens, C., R. Futami, L. Covelli, L. Domínguez-Escribá, J. M. Viu, D. Tamarit, J. Aguilar-Rodríguez, M. Vicente-Ripolles, G. Fuster, and G. P. Bernet. 2010. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res* 39:D70–D74.

Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.

Mank, J. E. 2012. Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. *Chromosom Res* 20:21–33.

Mank, J. E., D. J. Hosken, and N. Wedell. 2014. Conflict on the sex chromosomes: cause, effect, and complexity. *Cold Spring Harb Perspect Biol* 6:a017715.

Marais, G. A. B., P. R. A. Campos, and I. Gordo. 2010. Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biol Evol* 2:347–357.

Marais, G. A. B., J.-M. Gaillard, C. Vieira, I. Plotton, D. Sanlaville, F. Gueyffier, and J.-F. Lemaitre. 2018. Sex gap in aging and longevity: can sex chromosomes play a role? *Biol Sex Differ* 9:1–14.

Martens, M., A. Ammar, A. Riutta, A. Waagmeester, D. N. Slenter, K. Hanspers, R. A. Miller, D. Digles, E. N. Lopes, and F. Ehrhart. 2021. WikiPathways: connecting communities. *Nucleic Acids Res* 49:D613–D621.

Martin, E. S., R. Cesari, F. Pentimalli, K. Yoder, R. Fishel, A. L. Himelstein, S. E. Martin, A. K. Godwin, M. Negrini, and C. M. Croce. 2003. The BCSC-1 locus at chromosome 11q23-q24 is a candidate tumor suppressor gene. *PNAS* 100:11517–11522.

Matsubara, K., Y. Kumazawa, H. Ota, C. Nishida, and Y. Matsuda. 2019. Karyotype analysis of four blind snake species (Reptilia: Squamata: Scolecophidia) and karyotypic changes in Serpentes. *Cytogenet Genome Res* 157:98–106.

Matsubara, K., H. Tarui, M. Toriba, K. Yamada, C. Nishida-Umehara, K. Agata, and Y. Matsuda. 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *PNAS* 103:18190–18195.

Meiklejohn, C. D., and Y. Tao. 2010. Genetic conflict and sex chromosome evolution. *Trends Ecol Evol* 25:215–223.

Mi, H., D. Ebert, A. Muruganujan, C. Mills, L.-P. Albou, T. Mushayamaha, and P. D. Thomas. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res* 49:D394–D403.

Mi, H., A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo, and P. D. Thomas. 2019. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v. 14.0). *Nat Protoc* 14:703–721.

Mi, H., and P. Thomas. 2009. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. In *Protein Networks and Pathway Analysis*, pp 123–140. Y. Nikolsky and J. Bryant eds. Humana Press.

Mistry, J., S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, and L. J. Richardson. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* 49:D412–D419.

Morchikh, M., A. Cribier, R. Raffel, S. Amraoui, J. Cau, D. Severac, E. Dubois, O. Schwartz, Y. Bennasser, and M. Benkirane. 2017. HEXIM1 and NEAT1 long non-coding RNA form a multi-subunit complex that regulates DNA-mediated innate immune response. *Mol Cell* 67:387–399.

Nam, K., and H. Ellegren. 2008. The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics* 180:1131–1136.

Nguyen, A. H., and D. Bachtrog. 2020. Toxic Y chromosome: increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome. bioRxiv.

Nguyen, A. H., and D. Bachtrog. 2021. Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male *Drosophila* with a young Y chromosome. *PLoS Genet* 17:e1009438.

Nilsson, S., S. Makela, E. Treuter, M. Tujague, J. Thomsen, G. Andersson, E. Enmark, K. Pettersson, M. Warner, and J.-Å. Gustafsson. 2001. Mechanisms of estrogen action. *Physiol Rev* 81:1535–1565.

O'Meally, D., T. Ezaz, A. Georges, S. D. Sarre, and J. A. M. Graves. 2012. Are some chromosomes particularly good at sex? Insights from amniotes. *Chromosom Res* 20:7–19.

Ohno, S. 1967. Sex chromosomes and sex-linked genes. Springer, Berlin.

Parsch, J., and H. Ellegren. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* 14:83–87.

Pasquesi, G. I. M., R. H. Adams, D. C. Card, D. R. Schield, A. B. Corbin, B. W. Perry, J. Reyes-Velasco, R. P. Ruggiero, M. W. Vandewege, J. A. Shortt, and T. A. Castoe. 2018. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun* 9:1–11.

Pedersen, B. S., and A. R. Quinlan. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34:867–868.

Peona, V., O. M. Palacios-Gimenez, J. Blommaert, J. Liu, T. Haryoko, K. A. Jønsson, M. Irestedt, Q. Zhou, P. Jern, and A. Suh. 2021. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. *Philos Trans R Soc B* 376:20200186.

Perry, B. W., D. C. Card, J. W. McGlothlin, G. I. M. Pasquesi, R. H. Adams, D. R. Schield, N. R. Hales, A. B. Corbin, J. P. Demuth, F. G. Hoffmann, M. W. Vandewege, R. K. Schott, N. Bhattacharyya, B. S. W. Chang, N. R. Casewell, G. Whiteley, J. Reyes-Velasco, S. P. Mackessy, T. Gamble, K. B. Storey, K. K. Biggar, C. N. Passow, C.-H. Kuo, S. E. McGaugh, A. M. Bronikowski, A. P. J. de Koning, S. V Edwards, M. E. Pfrender, P. Minx, E. D. Brodie, E. D. Brodie, W. C. Warren, and T. A. Castoe. 2018. Molecular adaptations for sensing and securing prey and insight into amniote genome diversity from the garter snake genome. *Genome Biol Evol* 10:2110–2129.

Piergentili, R. 2010. Multiple roles of the Y chromosome in the biology of *Drosophila melanogaster*. *ScientificWorldJournal* 10:1749–1767.

Rice, E. S., S. Kohno, J. S. John, S. Pham, J. Howard, L. F. Lareau, B. L. O'Connell, G. Hickey, J. Armstrong, A. Deran, I. Fiddes, R. N. Platt 2nd, C. Gresham, F. McCarthy, C. Kern, D. Haan, T. Phan, C. Schmidt, J. R. Sanford, D. A. Ray, B. Paten, L. J. Guillette Jr., and R. E. Green. 2017. Improved genome assembly of American alligator genome reveals conserved architecture of estrogen signaling. *Genome Res* 27:686–696.

Rogers, T. F., T. Pizzari, and A. E. Wright. 2021. Multi-copy gene family evolution on the avian W chromosome. *J Hered* 112:250–259.

Rovatsos, M., M. Altmanova, M. Johnson Pokorná, B. Augstenova, and L. Kratochvíl. 2018. Cytogenetics of the Javan file snake (*Acrochordus javanicus*) and the evolution of snake sex chromosomes. *J. Zool Syst Evol Res* 56:117–125.

Rovatsos, M., J. Vukić, P. Lymberakis, and L. Kratochvíl. 2015. Evolutionary stability of sex chromosomes in snakes. *Proc R Soc B* 282:20151992.

Rozen, S., H. Skaletsky, J. D. Marszalek, P. J. Minx, H. S. Cordum, R. H. Waterston, R. K. Wilson, and D. C. Page. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423:873–876.

Scherdin, U., K. Rhodes, and M. Breindl. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* 64:907–912.

Schield, D. R., D. C. Card, N. R. Hales, B. W. Perry, G. I. M. Pasquesi, H. Blackmon, R. H. Adams, A. B. Corbin, C. F. Smith, B. Ramesh, J. P. Demuth, E. Betran, M. Tollis, J. M. Meik, S. P. Mackessy, and T. A. Castoe. 2019. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res* 29:590–601.

Schield, D. R., B. W. Perry, R. H. Adams, M. L. Holding, Z. L. Nikolakis, S. S. Gopalan, C. F. Smith, J. M. Parker, J. M. Meik, M. DeGiorgio, S. P. Mackessy, and T. A. Castoe. 2022. The roles of balancing selection and recombination in the evolution of rattlesnake venom. *Nat Ecol Evol*.

Schield, D. R., B. W. Perry, Z. L. Nikolakis, S. P. Mackessy, and T. A. Castoe. 2021. Population genomic analyses confirm male-biased mutation rates in snakes. *J Hered* 112:221–227.

Schröder, A. R. W., P. Shinn, H. Chen, C. Berry, J. R. Ecker, and F. Bushman. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110:521–529.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, and J. Söding. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.

Singchat, W., S. F. Ahmad, N. Laopichienpong, A. Suntronpong, T. Panthum, D. K. Griffin, and K. Srikulnath. 2020. Snake W sex chromosome: The shadow of ancestral amniote super-sex chromosome. *Cells* 9:2386.

Slattery, J. P., L. Sanner-Wachter, and S. J. O'Brien. 2000. Novel gene conversion between XY homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). *PNAS* 97:5307–5312.

Smeds, L., V. Warmuth, P. Bolivar, S. Uebbing, R. Burri, A. Suh, A. Nater, S. Bureš, L. Z. Garamszegi,

and S. Hogner. 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nat Commun* 6:7330.

Smit, A. F. A., R. Hubley, and P. Green. 2015. RepeatMasker Open-4.0. 2013–2015. *Inst Syst Biol* http//repeatmasker. org.

Soh, Y. Q. S., J. Alföldi, T. Pyntikova, L. G. Brown, T. Graves, P. J. Minx, R. S. Fulton, C. Kremitzki, N. Koutseva, and J. L. Mueller. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* 159:800–813.

Stanke, M., and B. Morgenstern. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33:W465–W467.

Steinbiss, S., U. Willhoeft, G. Gremme, and S. Kurtz. 2009. Fine-grained annotation and classification of *de novo* predicted LTR retrotransposons. *Nucleic Acids Res* 37:7002–7013.

Steinemann, M., and S. Steinemann. 1992. Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. *PNAS* 89:7591–7595.

Sultanova, Z., P. A. Downing, and P. Carazo. 2020. Genetic sex determination and sex-specific lifespan in tetrapods–evidence of a toxic Y effect. *bioRxiv*.

Suryamohan, K., S. P. Krishnankutty, J. Guillory, M. Jevit, M. S. Schröder, M. Wu, B. Kuriakose, O. K. Mathew, R. C. Perumal, I. Koludarov, L. D. Goldstein, K. Senger, M. Davis Dixon, D. Velayutham, D. Vargas, S. Chaudhuri, M. Muraleedharan, R. Goel, Y. J. Chen, A. Ratan, P. Liu, B. Faherty, G. de la Rosa, H. Shibata, M. Baca, M. Sagolla, J. Ziai, G. A. Wright, D. Vucic, S. Mohan, A. Antony, J. Stinson, D. S. Kirkpatrick, R. N. Hannoush, S. Durinck, Z. Modrusan, E. W. Stawiski, K. Wiley, T. Raudsepp, R. Manjunatha Kini, A. Zachariah, and S. Seshagiri. 2020. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat Genet* 52:106–117.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612.

R Core Team. 2017. R: A language and environment for statistical computing.

Tomaszkiewicz, M., P. Medvedev, and K. D. Makova. 2017. Y and W chromosome assemblies: approaches and discoveries. *Trends Genet* 33:266–282.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A.

Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. 2001. The sequence of the human genome. *Science* 291:1304–1351.

Vicoso, B., J. J. Emerson, Y. Zektser, S. Mahajan, and D. Bachtrog. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol* 11:e1001643.

Vonk, F. J., N. R. Casewell, C. V Henkel, A. M. Heimberg, H. J. Jansen, R. J. R. McCleary, H. M. E. Kerkkamp, R. A. Vos, I. Guerreiro, J. J. Calvete, W. Wuster, A. E. Woods, J. M. Logan, R. A. Harrison, T. A. Castoe, A. P. J. de Koning, D. D. Pollock, M. Yandell, D. Calderon, C. Renjifo, R. B. Currier, D. Salgado, D. Pla, L. Sanz, A. S. Hyder, J. M. C. Ribeiro, J. W. Arntzen, G. E. E. J. M. van den Thillart, M. Boetzer, W. Pirovano, R. P. Dirks, H. P. Spaink, D. Duboule, E. McGlinn, R. M. Kini, and M. K. Richardson. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *PNAS* 110:20651–20656.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Kunstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V Mello, P. Minx, P. Lovell, T. A. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y. C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backstrom, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Volker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V Edwards, S. P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, and R. K. Wilson. 2010. The genome of a songbird. *Nature* 464:757–762.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, and P. An. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.

Wei, K. H.-C., L. Gibilisco, and D. Bachtrog. 2020. Epigenetic conflict on a degenerating Y chromosome increases mutational burden in *Drosophila* males. *Nat Commun* 11:1–9.

Weisenfeld, N. I., V. Kumar, P. Shah, D. M. Church, and D. B. Jaffe. 2017. Direct determination of diploid genome sequences. *Genome Res* 27:757–767.

Wright, A. E., R. Dean, F. Zimmer, and J. E. Mank. 2016. How to make a sex chromosome. *Nat Commun* 7:1–8.

Wright, A. E., P. W. Harrison, S. H. Montgomery, M. A. Pointer, and J. E. Mank. 2014. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution* 68:3281–3295.

Wu, X., Y. Li, B. Crise, and S. M. Burgess. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300:1749–1751.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.

Yin, W., Z. J. Wang, Q. Y. Li, J. M. Lian, Y. Zhou, B. Z. Lu, L. J. Jin, P. X. Qiu, P. Zhang, W. B. Zhu, B. Wen, Y. J. Huang, Z. L. Lin, B. T. Qiu, X. W. Su, H. M. Yang, G. J. Zhang, G. M. Yan, and Q. Zhou. 2016. Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat Commun* 7:13107.

Zaher, H., R. W. Murphy, J. C. Arredondo, R. Graboski, P. R. Machado-Filho, K. Mahlow, G. G. Montingelli, A. B. Quadros, N. L. Orlov, and M. Wilkinson. 2019. Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). *PLoS One* 14:e0216148.

Zhang, S. V, L. Zhuo, and M. W. Hahn. 2016. AGOUTI: improving genome assembly and annotation using transcriptome data. *Gigascience* 5:s13742-016.

Zheng, Y., and J. J. Wiens. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Mol Phylogenet Evol* 94:537–547.

Tables

Table 1. W chromosome assembly and annotation statistics.

Assembly	10X Supernova	10X Supernova + Agouti
Number of scaffolds	2,139	2,027
Scaffold N50 (bp)	12,080	13,252
Longest scaffold (bp)	62,716	153,158
Annotated genes	213	219

Table 2. Refugium index (RI) for transposable element classes on the autosomes, Z chromosome, and W chromosome. χ^2 tests were used to test for deviations from uniform expectation.

Element	Autosome RI	Z chromosome RI	W chromosome RI	χ^2
Total repeats	-4.375	21.439	107.815	14,226,220***
ERVs	-42.164	69.973	1664.497	8,724,788***
mdg4	-24.472	87.749	750.403	17,352,014***
L1 LINEs	-16.849	70.555	470.236	6,583,173***
CR1 LINEs	-0.6362	5.9233	2.842	14,663***
fl-LTRs	-4.09	28.895	60.402	1,604,306***

*** p-value < 2.2×10^{-16}; ERVs = endogenous retroviruses; fl-LTRs = full-length LTR retrotransposons.

Figure Captions

Figure 1. Genetic sex determination in squamates, homology between prairie rattlesnake Z and W chromosomes and the presence of evolutionary strata. **A.** Phylogeny of representative snake species and *Anolis* outgroup with known (or presumed) genetic sex determination systems. Henophidian species are highlighted in grey and caenophidian species are highlighted in blue. The bracketed arrow indicates species in the Colubriformes subclade of Caenophidia. Branches are scaled by divergence times in millions of years. **B.** Locations of homologous W chromosome scaffolds (blue) and annotated coding genes (i.e., ZW gametologs; orange) on the Z chromosome. The differently shaded sections of the Z chromosome ideogram represent distinct regions identified in Schield et al. (2019); older strata (green), RS = recent stratum (turquoise), PAR = pseudoautosomal region (grey). The approximate location of the centromere is shown in grey. **C.** Synonymous divergence (d_S) between Z and W gametologs anchored to Z chromosome positions. The three shaded regions correspond to spatial clusters of gametologs with similar levels of divergence. **D.** Distributions of d_S by cluster with assignment to hypothesized evolutionary strata. **E.** Schematic of the hypothesized order of recombination suppression events among evolutionary strata between the Z and W chromosomes. The approximate dates for each event are based on mean divergence estimates per stratum.

Figure 2. Relative female coverage across the Z chromosome in four colubroid snake species, including a representative colubrid (western terrestrial garter snake) and three pitvipers (five-pace viper, pygmy rattlesnake, and prairie rattlesnake). The phylogeny and divergence times for the four species and *Python* are shown to the left, with the grey box denoting pitviper species. Shaded boxes depict approximate dates for the formation of the older and recent strata. Panels to the right for each species show variation in the normalized ratio of female:male sequencing read coverage ($\log_2$FM) in 10 kb sliding windows. Horizontal dashed lines at 0 and -1 correspond to expectations for autosomal and Z-linked regions in females, respectively, and the pseudoautosomal region (PAR) and evolutionary strata (RS = recent stratum) are shaded according to the above Z chromosome ideogram. Distributions of $\log_2$FM in each region are summarized at the far right; points represent median $\log_2$FM and whiskers represent Q1-1.5*IQR and Q3+1.5*IQR. Images were modified from photographs from: Steve Jurvetson (garter snake), Alex White (five-pace viper), Peter Paplanus (pygmy rattlesnake), and Blair Perry (prairie rattlesnake) under Creative Commons license CC BY-NC-SA 2.0.

Figure 3. Composition of the W chromosome and comparison with other amniotes. **A**. Distributions of GC content on autosomes, Z(X), and W(Y) chromosomes of the prairie rattlesnake and bird and mammal species. **B**. Distributions of CpG content on autosomes, Z(X), and W(Y) chromosomes. **C.** Age distribution of TEs on the W chromosome. Shaded bars represent the proportions of TE families with

increasing age (i.e., % Kimura substitution level). Mdg4 elements are denoted with a '*'. **D**. Proportions of autosomes, Z, and W chromosomes annotated as repeats. **E**. Proportions of autosomes, Z, and W chromosomes annotated as mdg4 LTR elements. **F**. Positive relationship between percent mdg4 elements and GC content across W chromosome scaffolds. **G**. No correlation between percent total repeats (excluding mdg4 elements) and GC content across W chromosome scaffolds.

Figure 4. Refugium indexes for the autosomes, Z, and W chromosomes for total repeats (**A**) and ERV (**B**), mdg4 (**C**), L1 LINE (**D**), and CR1 LINE (**E**) retroelements, calculated based on observed an expected frequencies of repeat bp on the autosomes and sex chromosomes, assuming a uniform distribution. Bar plots in the top of each panel show calculated refugium index values; refugium index > 0 indicates an excess of repeats; refugium index < 0 indicates depletion of repeats relative to a uniform distribution. All autosomal refugium indices are negative. Pie charts at the bottom of each panel show the proportions of autosomal, Z, and W chromosome bp in the analyzed genome assembly compared to proportions of bp for each repeat element category.

Figure 5. Gene expression and functional characterization of genes on the W chromosome. **A**. Log_{10}-normalized TPM expression counts for ovary, kidney, and liver in females are compared to testes, kidney, and liver in males. Female and male average expression values are displayed in the center. Bars at the right denote genes within the recent and older evolutionary strata. Genes with female-biased expression are denoted with red dots to the left of the heatmap. **B**. Gene ontology categories (biological processes and cellular components) that comprise W-linked genes. Terms denoted with asterisks are significantly enriched for W-linked genes after FDR correction. **C**. Panther protein classification for W-linked genes.

Figure 6. W-specific gene duplications, sex-linked divergence between pitviper species, and GC content at third codon positions (GC3) in rattlesnake ZW gametologs. **A**. Estimated copy number for W-linked genes derived from comparisons with autosomal gene coverage, with points placed according to their position on the Z chromosome. The shaded regions correspond to strata 1 (older) and 2 (recent). Note that the y-axis is not continuous to display genes with high estimated copy number. Points for the genes with evidence of two or more copies are labeled. Gene names appearing more than once had multiple annotated positions on W-linked scaffolds. Genes without an identified 1:1 Z gametolog are labeled 'UN' on the x-axis. **B**. Distributions of d_N/d_S between prairie rattlesnake and five-pace viper Z and W orthologs in stratum 1 and 2. **C**. Distributions d_S between prairie rattlesnake Z and W orthologs. **D**. GC3 content of ZW gametologs in prairie rattlesnake. The scatterplot to the left shows the relationship between GC3 for Z and W gametologs. The panels to the right show GC3 content for W and Z gametologs, respectively, based on Z chromosome position.

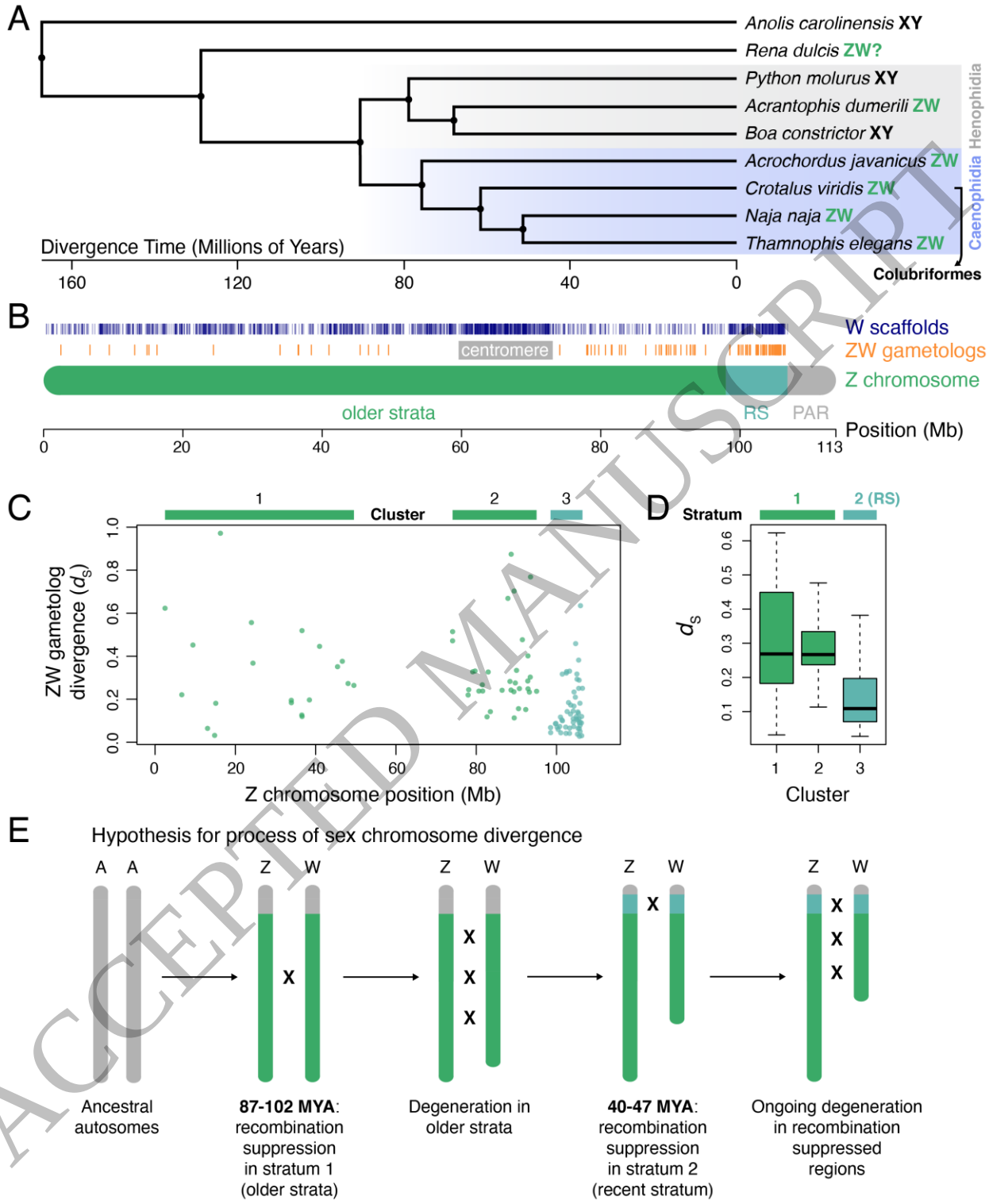


Figure 1
165x205 mm (.57 x DPI)

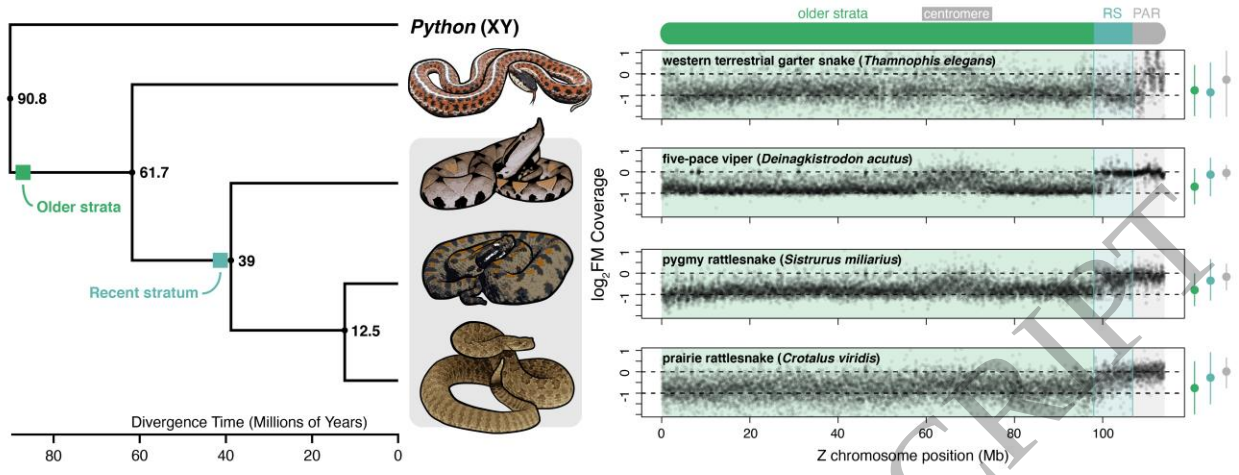


Figure 2
165x62 mm (.57 x DPI)

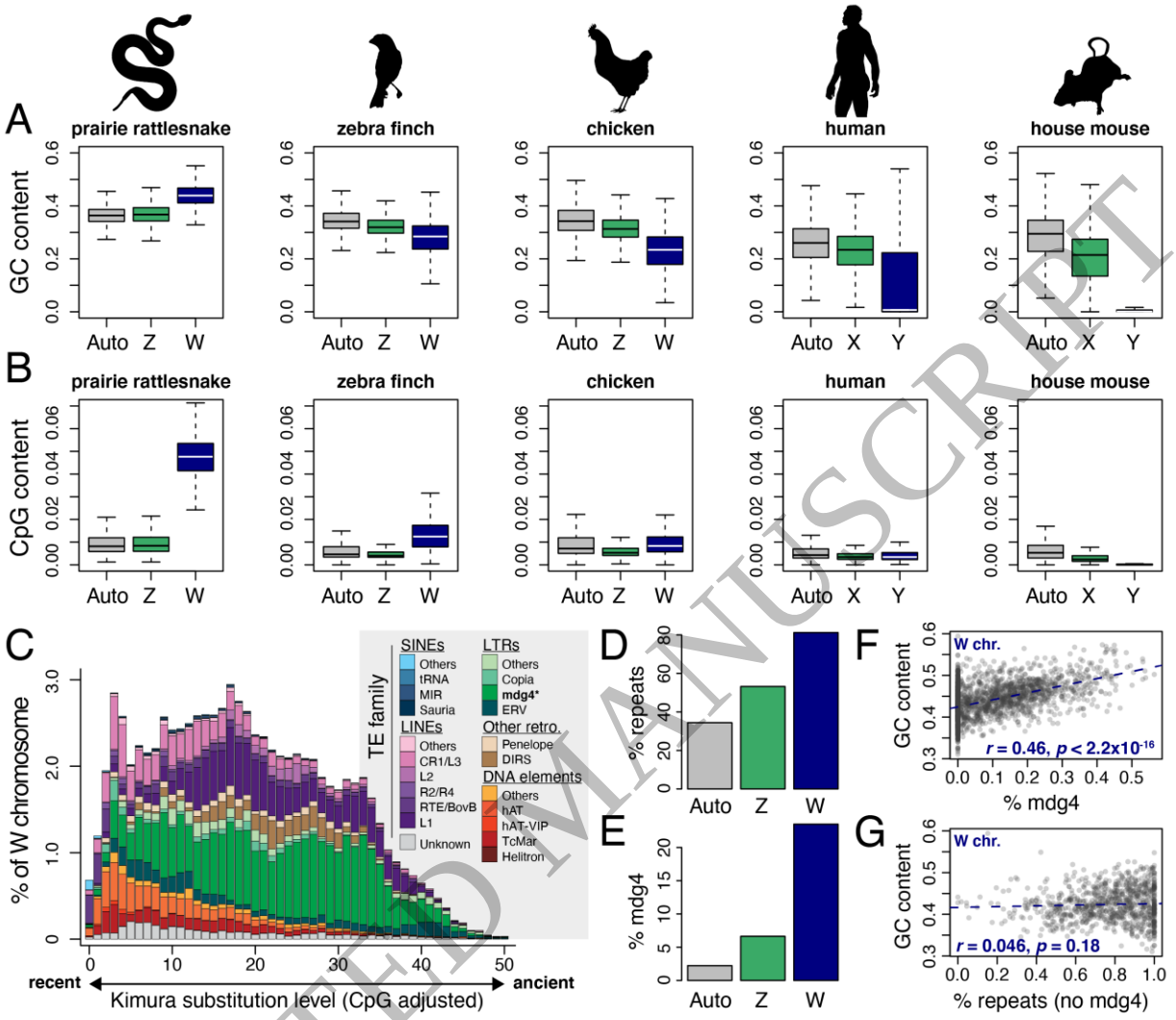
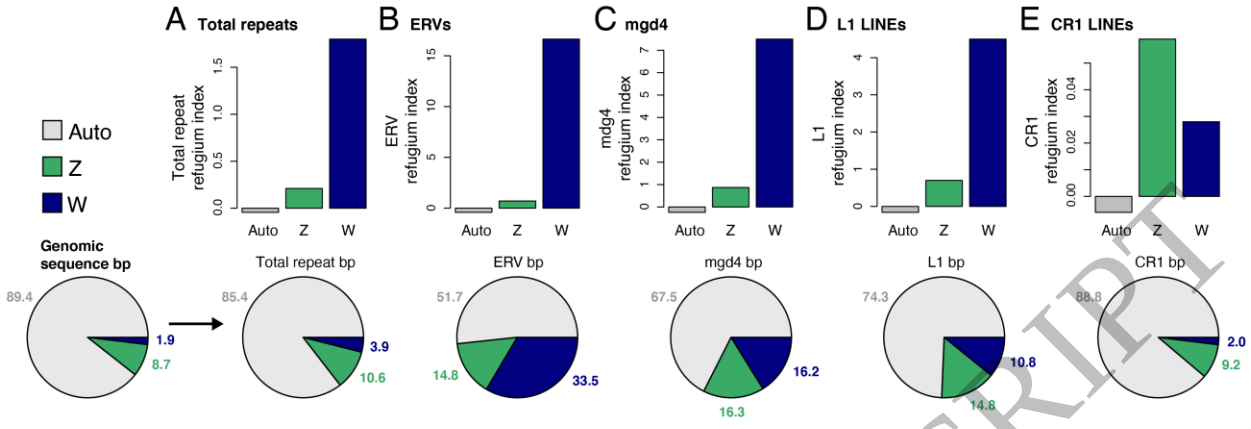


Figure 3
165x142 mm (.57 x DPI)



1

2 **Figure 4**

3 **165x56 mm (.57 x DPI)**

4

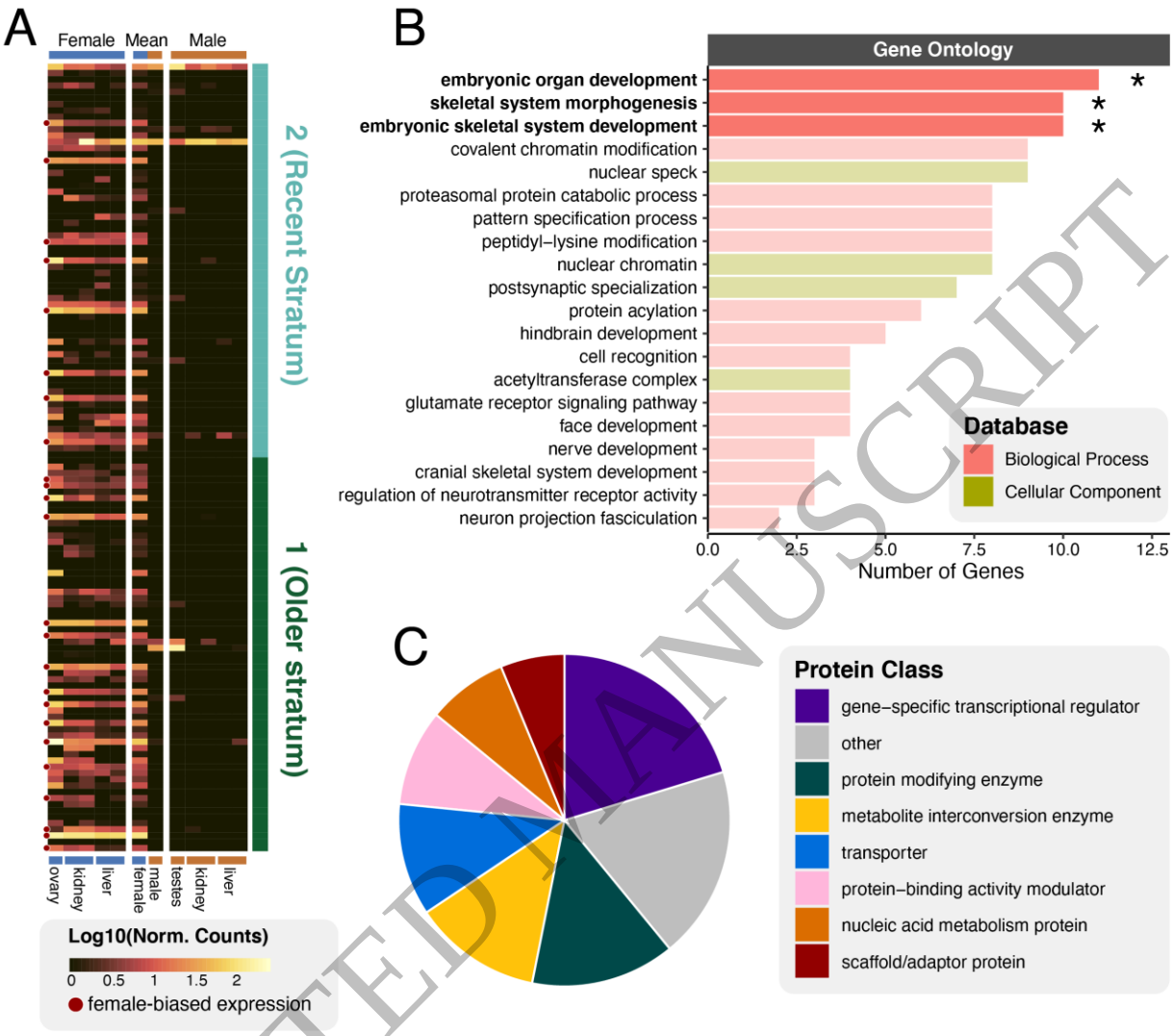


Figure 5
165x144 mm (.57 x DPI)

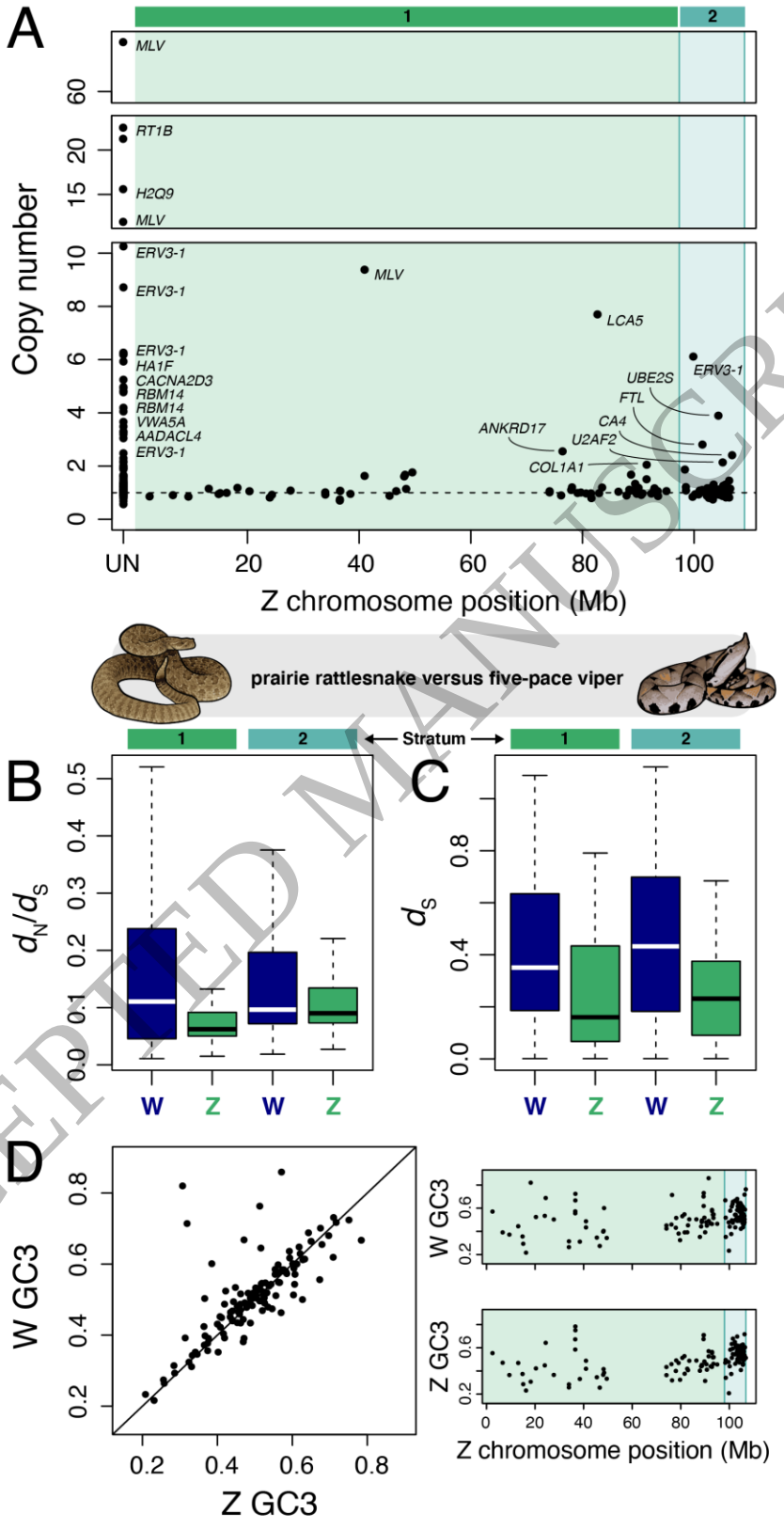


Figure 6
117x229 mm (.57 x DPI)