

Sex chromosome identification and genome curation from a single individual with SCINKD

Brendan J. Pinto  *^{1,2,3†}, Simone M. Gable  ^{4†}, Shannon E. Keating  ^{5†}, Chase H. Smith  ^{6†},
Tony Gamble  ^{2,5,7}, Stuart V. Nielsen  ^{8,9†}, Melissa A. Wilson  *^{1,3}

¹Comparative Genomics and Reproductive Health Section, Center for Genomics and Data Science Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

²Department of Zoology, Milwaukee Public Museum, Milwaukee, WI, USA

³Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA

⁴Global Futures Laboratory, Arizona State University, Tempe, AZ, USA

⁵Department of Biological Sciences, Marquette University, Milwaukee, WI, USA

⁶Department of Integrative Biology, University of Texas, Austin, TX, USA

⁷Bell Museum of Natural History, University of Minnesota, St. Paul, MN, USA

⁸Museum of Life Sciences, Department of Biological Sciences, Louisiana State University Shreveport, Shreveport, LA, USA

⁹Department of Natural History, Florida Museum of Natural History, Gainesville, FL, USA

*Corresponding authors: E-mails: brendanjohnpinto@gmail.com; melissa.wilson3@nih.gov.

†Co-first author.

‡Former affiliation.

Associate editor: Amanda Larracuent

Abstract

In most animal species, the sex-determining pathway is typically initiated by the presence/absence of a primary genetic cue at a critical point during development. This primary genetic cue is often located on a single locus—referred to as sex chromosomes—and can be limited to females (in a ZZ/ZW system) or males (in an XX/XY system). One trademark of sex chromosomes is a restriction or cessation of recombination surrounding the sex-limited region (to prevent its inheritance in the homogametic sex). This may lead to—through a variety of mechanisms—higher amounts of genetic divergence within this region, ie between the X/Z and Y/W chromosomes, especially when compared with their autosomal counterparts. Recent advances in genome sequencing and computation have brought with them the ability to resolve haplotypes within a diploid individual, permitting assembly of previously challenging genomic regions like sex chromosomes. Leveraging these advances, we identified replicable diagnostic characteristics between typical autosomes and sex chromosomes (within a single genome assembly). Under this framework, we can use this information to identify putative sex chromosome linkage groups across divergent vertebrate taxa and simultaneously curate misassembled regions on autosomes. Here, we present this conceptual framework and associated tool for identifying candidate sex chromosome linkage groups from a single, diploid individual dubbed Sex Chromosome Identification by Negating Kmer Densities.

Keywords sex chromosomes, genomics, genome curation, kmer-based, biodiversity genomics

Introduction

In species that reproduce sexually, populations include individuals that produce eggs (females) or sperm (males). The process of sex determination can be initiated via environmental or genetic cues. In diploid vertebrates with environmental sex determination, all individuals possess two copies of each chromosome that can recombine over their entire length. By contrast, most vertebrate species possess sex chromosomes with genetic differences that lead to a developmental cascade becoming typical females and males, ie genetic sex determination. The locus containing a gene(s) of large effect in sex determination is often located in a region that has restricted recombination, or does not recombine, and can therefore be inherited either strictly maternally (W-specific in a ZZ/ZW

system) or paternally (Y-specific in a XX/XY system). Over time, these sex-limited regions are often subjected to forces that suppress recombination—whereby mutations accumulate quickly from a reduced efficiency of DNA repair due to the lack of homologous recombination in this region (Sansam and Pezza 2015; Raina et al. 2025). Indeed, we know that this mode of sex chromosome evolution is widespread in animals, if not the majority of systems (Marshall Graves 2008; Graves 2016; Jeffries et al. 2025). Developing methods to quickly identify candidate sex-linked regions in genome assemblies at scale is of utmost importance as biodiversity genomics data across the tree of life are now being produced at unprecedented, albeit unequal, rates across the globe (Rhie et al. 2021; Card et al. 2023; Gable et al. 2023; Pinto et al. 2023b; Linck and Cadena 2024).

Received: July 25, 2025. **Revised:** February 20, 2026. **Accepted:** February 25, 2026

Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution 2026.

This work is written by (a) US Government employee(s) and is in the public domain in the US.

Over the past ~15 years, empiricists have developed many techniques to identify sex chromosomes using genetic sequence data from multiple female and male samples (Palmer et al. 2019). In the absence of diverse panels of reference-quality genomes for nonmodel organisms, the revolution in sex chromosome identification began using short read-based genomic sequencing with RADseq (Gamble and Zarkower 2014; Feron et al. 2021). Within the last decade, the progression toward whole-genome sequencing (WGS) data was accompanied by the requirement of a reference genome (Conte et al. 2017; Pinto et al. 2024). Most recently, the rise of the use of kmers in WGS data have pushed us back toward a reference-optional approach to identifying sex chromosomes (Behrens et al. 2024; Carey et al. 2024). However, with the advent of relatively inexpensive long read sequencing, generating diploid genome assemblies in biodiverse taxa has become more approachable, while the availability of population-level genomic sampling in many taxa has remained stagnant. Indeed, while often overlooked as sequencing has become easier and costs significantly decreased, field collections of population-level samples of plants, animals, and other eukaryotes have remained challenging. Thus, at its most extreme in this biodiversity genomics age, it is now possible to sequence and assemble the haplotype-resolved sex chromosome complement in a reference genome, but neglect to properly identify or annotate (formally label) them (see herein).

Prior DNA sequencing platforms (short-read and noisy long-read) were hindered by an inability to assemble high-quality genomes, especially the sex chromosomes, at scale. This issue has caused a historical, and intentional, bias toward sequencing mostly the homogametic sex (XX/ZZ individuals) in well-known taxa, such as mammals (XX), caenophidian snakes (ZZ), and birds (ZZ) to improve contiguity in de novo genome assemblies (Carvalho and Clark 2013). However, cutting-edge sequencing technologies and computational methods have brought unprecedented insight into genome form and function (Cheng et al. 2022; Antipov et al. 2025), permitting access to previously uncharacterized genomic regions, including sex chromosomes at scale (Makova et al. 2024; Rhie et al. 2023). However, characterization of sex chromosomes in nonmodel systems remains challenging and, again, currently relies on generating additional data beyond a reference genome assembly, which is prohibitive in many instances (Pinto et al. 2022, 2023b, 2024). To address the gap between reference genome availability and supporting data, we developed an approach to identify sex chromosomes from a single, diploid individual's genome (Fig. 1). This method is also useful for identifying finer-scale phasing and assembly issues. Leveraging a few modest assumptions regarding within-genome correlations, which appear to be conserved across most vertebrates, we describe Sex Chromosome Identification by Negating Kmer Densities (SCINKD) to identify unannotated sex chromosomes and curate diploid genome assemblies from a single individual.

Materials and methods

The SCINKD [v2.1] workflow is three steps, requiring as input a haplotype-resolved genome assembly using a haplotype-aware

assembler, such as hifiasm or verkko (Cheng et al. 2021, 2022; Antipov et al. 2025), optimally scaffolded to chromosome-level. The steps are:

1. Count kmers ($k = 28$) from both assembled haplotypes using meryl (Rhie et al. 2020)
2. Negate kmers that occur in both haplotype assemblies, keeping only kmers unique to each assembly.
3. Identify loci with kmer densities that deviate from the autosomal background using multivariate data visualization (dot plots, mirror plots, and haplotype alignments).

The tool to conduct these steps is developed as a snakemake workflow and template R files for data visualization; the FULL pipeline comes preoptimized for most genomes (<~5 Gb) and is co-distributed with a GREEDY mode that uses homopolymer compression. However, the GREEDY mode remains largely untested for this purpose due to lack of high-quality assemblies of large genomes; it has potential uses with larger genomes (>5 Gb). Both implementations come preconditioned for use on machines with 24 available threads and 24 Gb of RAM. The full implementation of this workflow and a detailed tutorial on how to run and input requirements is available on GitHub (<https://github.com/DrPintoThe2nd/SCINKD>).

Generation of new data

We extracted high molecular-weight (HMW) DNA from the liver of adult, male *Sphaerodactylus notatus* via SOP-CEPC (Pinto et al. 2021) and generated a HiC library from the same individual using a DoveTail Omni-C kit (Cantata Bio; Cambridge, MA, USA). A single PacBio HiFi SMRTbell library was constructed from the extracted HMW DNA, which was subsequently barcoded and sequenced across 1.5 SMRT cells at the Arizona Genomics Institute (AGI) (University of Arizona; Tucson, AZ, USA). We sequenced the HiC library on an Illumina NovaSeq 6000 at the Texas A&M Agrilife Core Facility (College Station, TX, USA). We interrogated sequence features using fastqc (Andrews 2010) and GenomeScope (Ranallo-Benavidez et al. 2020). All other data used in this study were published and/or downloaded from public access databases (Table S1).

Data analysis

Previous work showed extensive within-genome correlations using only a single reference genome (Pinto et al. 2023b). To corroborate whether a null expectation of the number of SNPs per chromosome as a function of chromosome length exists, we re-analyzed a small subset of human data. We downloaded data from Genotype-Tissue Expression (GTEx) project, approved for project #8834 for General Research Use in Genotype-Tissue Expression (GTEx) to M.A.W. We identified six individuals with an XX genotype in the GTEx data that matched specific search criteria for an independent project. These data were archived in the genome-aligned CRAM format, thus, we re-processed them with a modified GRCh38 reference to account for sex chromosome complement following previously described methods (Pinto et al. 2023c) and calculated numbers of biallelic SNPs using

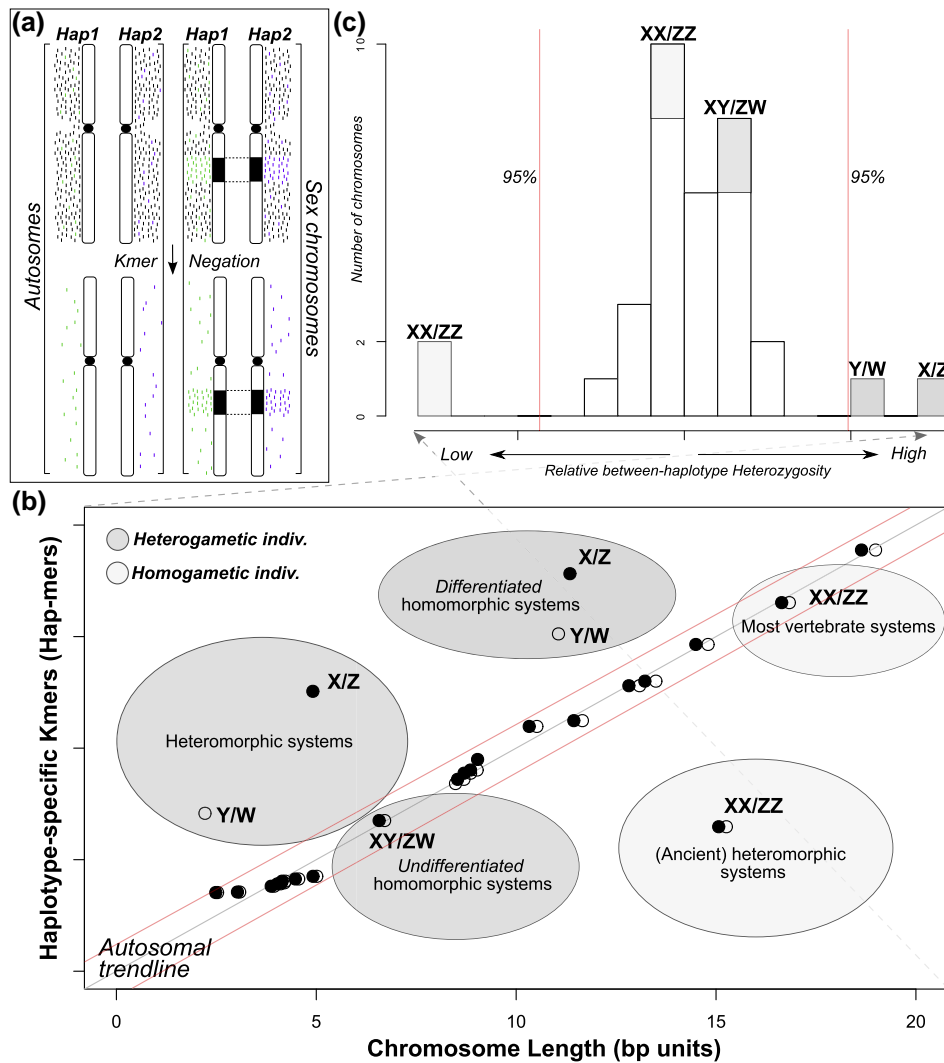


Figure 1 Simplified visual description of the SCINKD framework: (A) In many taxa, autosomes and sex chromosomes show disparate patterns where the region of restricted recombination (black) possesses increased haplotype-specific kmer (hap-mer) densities relative to the autosomal background. (B) This pattern drives overall increases in the number of hap-mers beyond what would be expected by chance (i.e. given the length of the chromosome). Each ‘bubble’ on the plot displays a hypothetical example of where the sex chromosomes of any given taxon could appear on the plot, including those that would be undetectable by SCINKD. (C) This deviation in hap-mers is detectable as a statistical outlier that can then be scrutinized using multivariate data visualization and/or additional data from samples of known sex.

rtg-tools (Cleary et al. 2015). We tested for significant differences between the number of SNPs across chr7 (159 Mb) and chr8 (145 Mb) versus chrX (156 Mb) using the Wilcoxon rank-sum test (Wilcoxon 1945).

We assembled HiFi reads with hifiasm [v0.24.0-r702] including HiC data for haplotype phasing (Cheng et al. 2021, 2022) for each test species: *Hemicordylus capensis* (Leitão et al. 2023), *Correlophus ciliatus* (Gumangan et al. 2024), *Eublepharis macularius* (Pinto et al. 2023a), *Lepidodactylus listeri* (Dodge et al. 2025), and *S. notatus* (“this study”) (Table S1). We scaffolded each haplotype separately using either juicebox [v1.6] (Durand et al. 2016) or Pretext (<https://github.com/sanger-tol/PretextMap>). We then conducted quality control of the assembly for gene content using BUSCO [v5.1.2] (Simão et al. 2015; Nishimura et al. 2017), calculated haplotype-specific statistics and visualizations by aligning the two haplotypes using minimap2 [v2.28-r1209] (Li

2018) and SVbyEye (Porubsky et al. 2025). For *Podarcis cretensis* (ZZ), we reassembled the diploid genome using hifiasm and scaffolded by aligning to the NCBI reference genome using RagTag (Alonge et al. 2022).

For each species with individuals of known sex, we used SCINKD to identify putative sex chromosomes within each genome and cross-validated with either the literature, additional data, or both (Pinto et al. 2022; Leitão et al. 2023). To test whether or not SCINKD would work in a species without HiC data, we reassembled the *Cryptoblepharus egeriae* genome, a species with higher quality data, but lacking HiC data (Dodge et al. 2025). Due to the limitations of assembling HiFi reads without HiC for phasing and scaffolding, we used the `--dual-scaf` setting in Hifiasm to generate comparable 1:1 contigs, and we then filtered contigs <1 Mb to reduce noise from genomic “shrapnel”. Lastly, for *S. notatus* and *C. ciliatus*, we leveraged recently developed

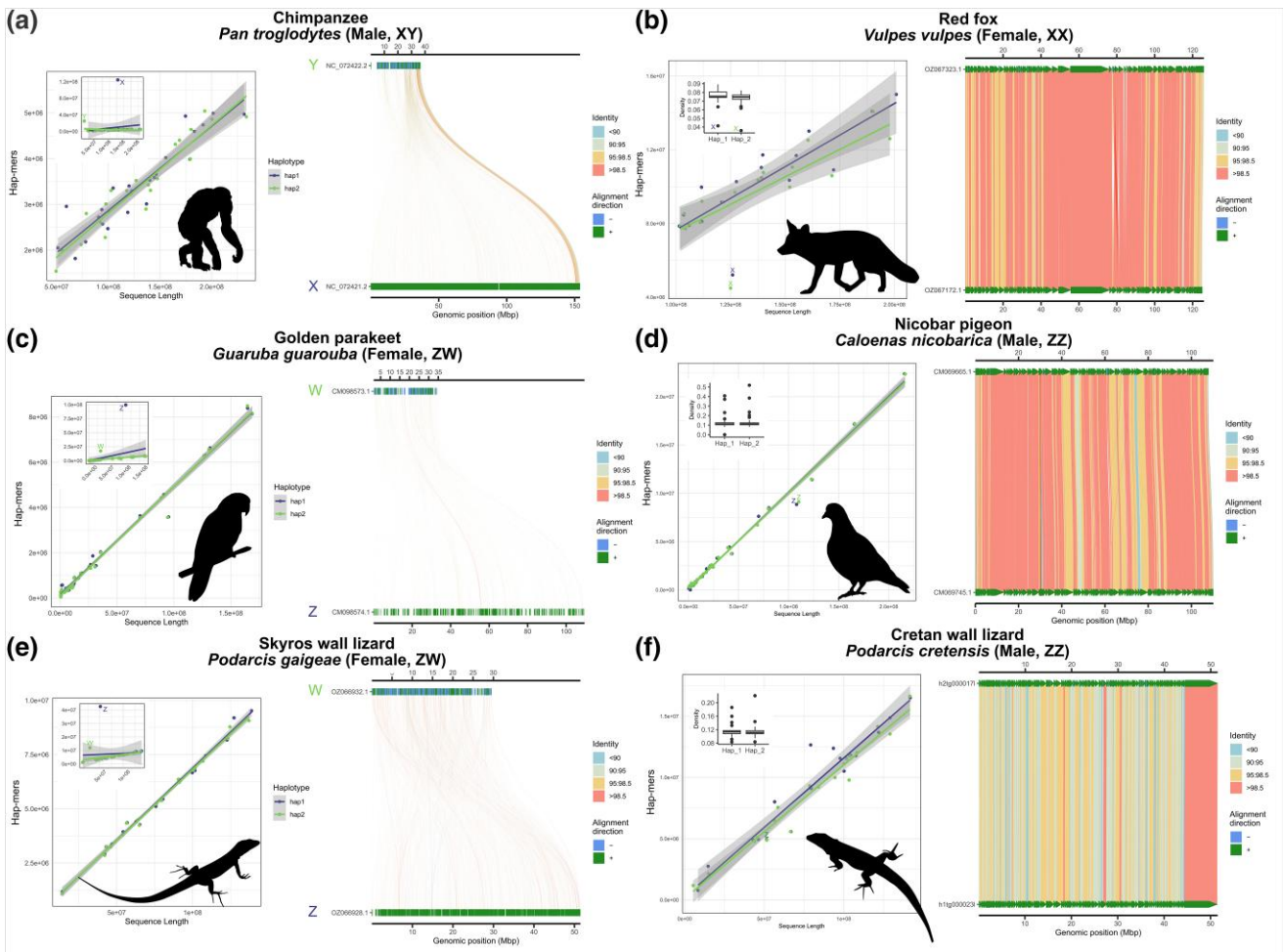


Figure 2 Exemplars of (A, B) mammal: XY (chimpanzee, *Pan troglodytes*) and XX (Red fox, *Vulpes vulpes*) individuals; (C, D) avian reptile: ZW (Golden parakeet, *Guaruba guarouba*) and ZZ (Nicobar pigeon, *Caloenas nicobarica*) individuals; (E, F) lacertid lizard: ZW (Skyros wall lizard, *Podarcis gaigeae*) and ZZ (Cretan wall lizard, *Podarcis cretensis*) individuals. Each panel consists of SCINKD “.results” output ~ sequence length (Li et al. 2009) visualized using ggplot2 (Wickham 2016) including only autosomes (exemplifying the predicted genome correlation), with an embedded panel demonstrating the full disparity between sex chromosomes and autosomes. Each panel also contains the corresponding minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025).

whole-genome data with previously published gametolog-specific RADtags to supplement SCINKD results. To confirm the sex chromosome system identified via SCINKD, we mapped these RADtags (X and Y or Z and W, respectively) to each of their respective diploid assemblies (Gamble et al. 2015; Nielsen et al. 2019; Keating 2022; Pinto et al. 2022; Gumangan et al. 2024).

To rigorously explore the sensitivity and features of SCINKD with regard to detecting misassemblies to flag for additional curation and the effects of sequencing depth and kmer length, we systematically interrogated the genome data of *Anniella stebbinsi* (a putative ZW female with a strong sex chromosome signal stemming from an ~10 Mb region at the distal end of chromosome 7) with two independent analyses. First, we re-assembled the haplotype-resolved reference genome using all available data (~47 \times coverage) as detailed above and systematically hard-masked a region on chromosome 2 in haplotype 1 of varying sizes (1, 5, 10, and 15 Mb) and ran SCINKD on each of the four modified haplotype pairs (Fig. S1). Next, we returned to the raw PacBio data and subsampled it at varying coverages (20 \times , 25 \times , 30 \times , and 40 \times)

using seqtk (<https://github.com/lh3/seqtk>). We assembled and scaffolded each of the eight haplotypes using HiFiasm and Pretext, respectively, as described previously. We then replicated SCINKD five times for each pair of haplotypes while varying kmer lengths: k21, k28 (default), k31, k41, and k51 (Fig. S2).

Results and Discussion

Recent work examining within-genome correlations sets strong expectations of linear relationships between chromosome length and content (GC content, gene density, SNPs, etc.) across vertebrates (Fig. S3) (Pinto et al. 2023b). In-line with this preliminary work, we hypothesized that this was also true for haplotype-specific kmers (hap-mers) in the leopard gecko (*E. macularius*), which is exactly what we found ($R^2 = 0.901$) (Fig. S4). Importantly, the leopard gecko only possesses autosomes, as it is a species with a well-supported temperature dependent sex determination system (TSD). This finding gave

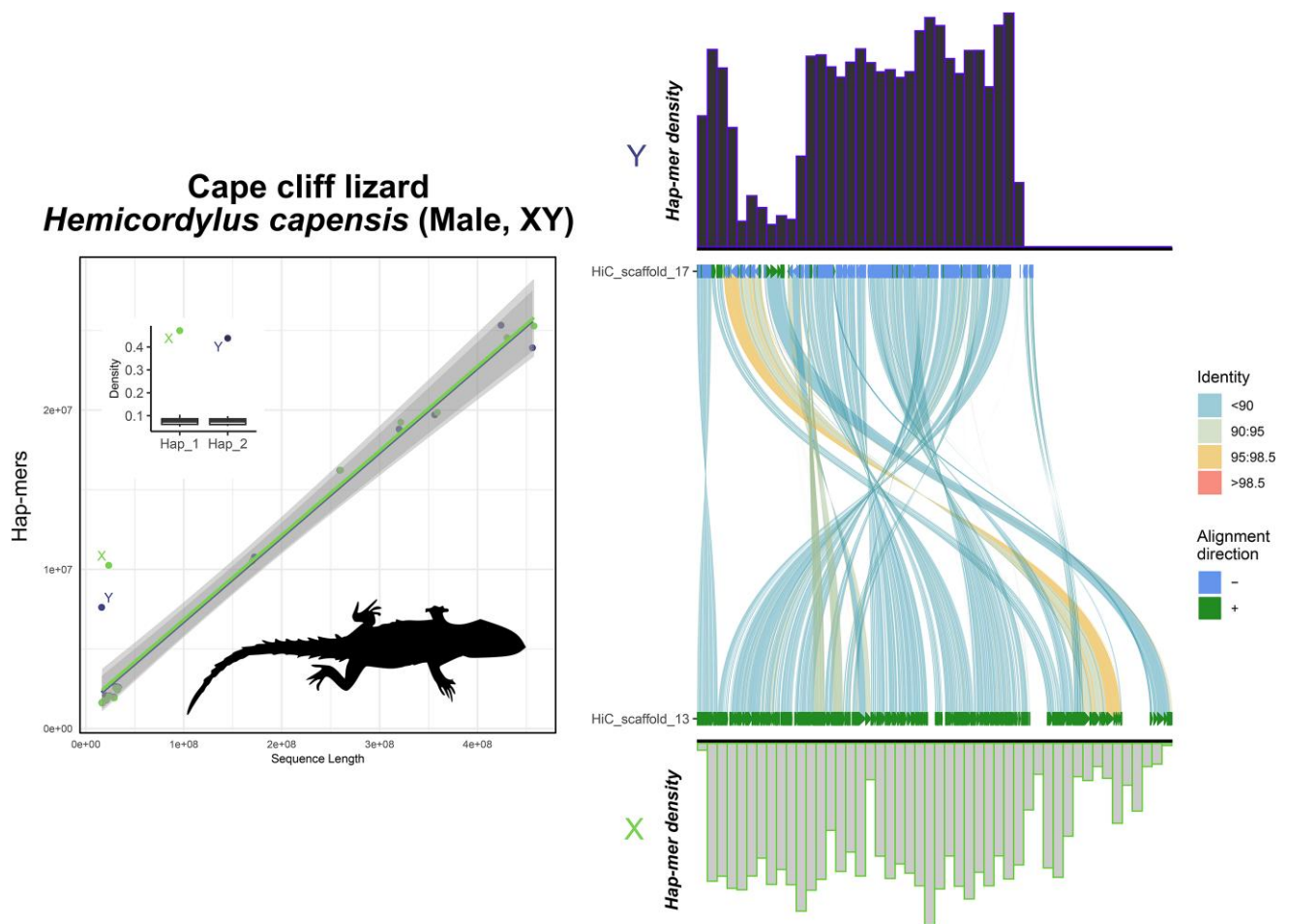


Figure 3 Total evidence plot for the Cape cliff lizard, *Hemicordylus capensis*, under the SCINKD framework. The left panel consists of SCINKD “results” output ~ sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1 Mb windows across the chromosome. A single pair of haplotypes deviates from the autosomal background. The signal comes from nearly the entirety of LG13, a homologous pair that also aligns with relatively low sequence identity across its length. There is no evidence that the PAR region is present in this assembly. The convergence of evidence on a single region corroborates previous work showing this is the XY linkage group in this scincomorph lizard species (Leitão et al. 2023).

rise to the prediction that hap-mers are correlated with chromosome length, except in regions that differ significantly in sequence composition, such as the nonrecombining region(s) of the sex chromosomes or severely misassembled genomic regions (Fig. 1a). To varying degrees, this is what we observe in our series of exemplar vertebrate datasets with known sex chromosome systems (Fig. 2). Thus, hap-mer densities are strongly correlated between homologous autosomes, but can be significantly different between the sex chromosomes across many taxa (Fig. 1).

Previously known exemplar cases

We developed a series of test cases from well-known XY and ZW systems from each of the three major amniote lineages ordered by relative species number (low to high): mammals, birds, and squamates (Fig. 2). Each has numerous high-quality genomic data, ranging from publicly available HiFi reads to complete T2T genomes. For each, we used the available GenBank assembly,

or reassembled from raw data when haplotype-resolved information was unavailable. In mammals, we chose the well-characterized, telomere-to-telomere (T2T) chimpanzee (*Pan troglodytes*; XY) and the chromosome-level red fox (*Vulpes vulpes*; XX) genomes. In both species, sex chromosomes are apparent when examining dot plots of SCINKD results by chromosome length (Fig. 2a and b). When aligning the two haplotypes for each species, we see the pseudoautosomal region (PAR) in chimpanzee aligns well, while the rest of the homologous pair aligns poorly. In the red fox, lower-than-expected heterozygosity on the pair of X chromosomes is exemplified by large stretches of high sequence similarity, replicating our expectations from human data (Fig. S3). In the two bird species, we see a similar pattern to chimpanzee in the Golden parakeet (ZW) and, to a much lesser extent, a similar pattern to the red fox in the Nicobar pigeon (ZZ) (Fig. 2c and d). Both therian mammal and avian sex chromosome systems are ancient, heteromorphic systems that generally serve as exceptional, not typical, vertebrate sex chromosome systems.

To balance these extreme examples, we examined well-characterized systems in a more diverse group, squamate

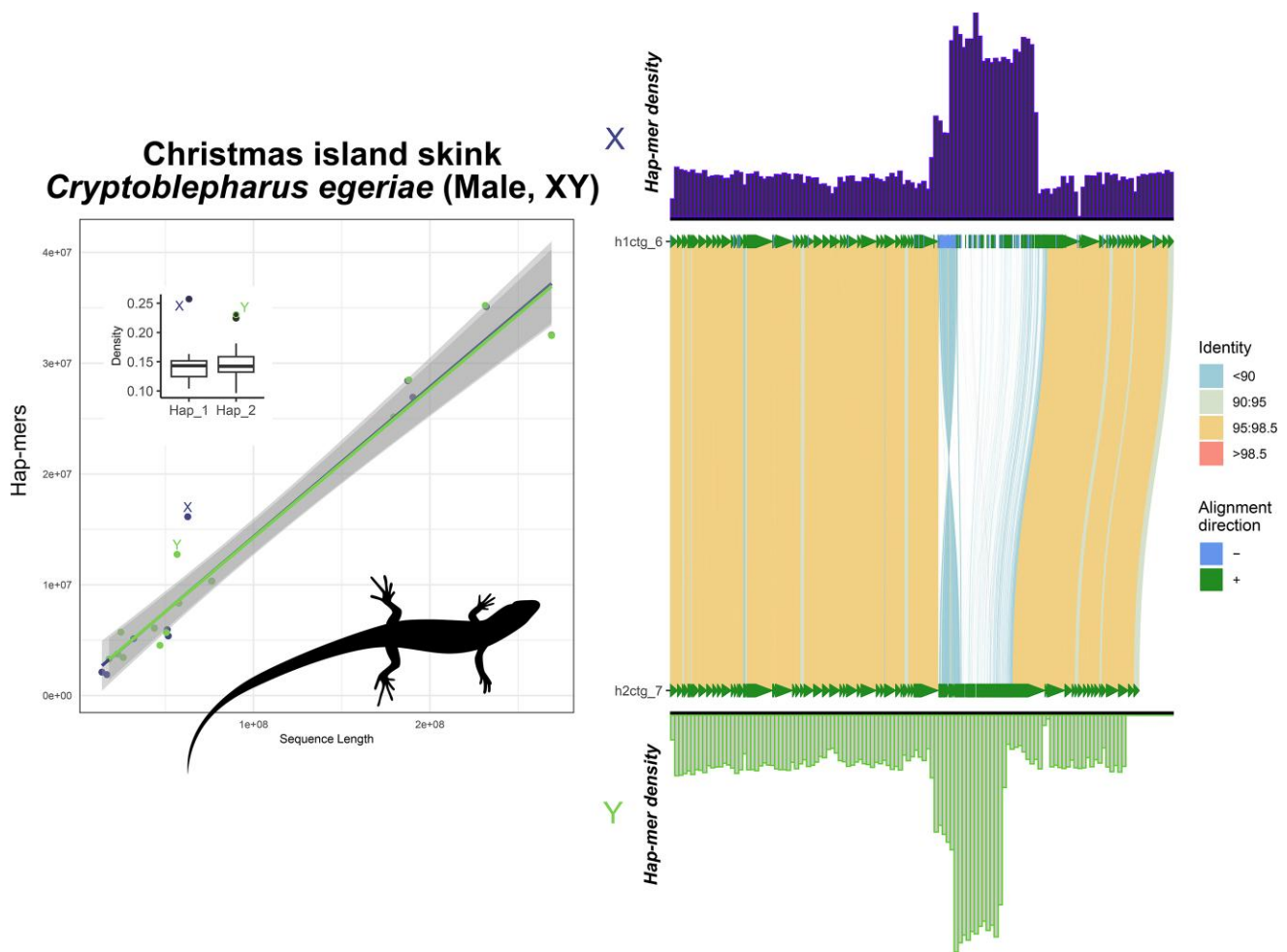


Figure 4 Total evidence plot for the Christmas Island skink, *Cryptoblepharus egeriae*, under the SCINKD framework. The left panel consists of SCINKD “results” output ~ sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1 Mb windows across the chromosome. A single pair of haplotypes deviates from the autosomal background. The signal itself originates from a single region at the center of contigs, a region that aligns poorly between haplotypes and possesses an excess of unique kmers. The convergence of evidence on a single region corroborates previous work in skinks and supports this region as the XY linkage group in this scincomorph lizard species (Kostmann et al. 2021) and was validated via alignment to chrX of the Three-lined skink, *Acritoscincus duperreyi*, assembly (GCA_041722995.2).

reptiles (Fig. 2e and f). In lizards of the genus *Podarcis* (wall lizards), we see the signs of their heteromorphic ZZ/ZW system, which is conserved at the family level (Lacertidae) (Rovatsos et al. 2016). In the Skyros wall lizard (*Podarcis gageae*; ZW), we observe a similar pattern to that of XY mammals and ZW birds, where the divergent nature of the heteromorphic system is apparent. However, in the Cretan wall lizard (*P. cretensis*; ZZ), the Z chromosomes show no visible diagnostic differences from any randomly selected pair of autosomes, which is a pattern expected for sex chromosome systems that are much more recently evolved than that of mammals or birds, and thus have not been subjected to the same kind of reduction in genetic diversity on the Z (Fig. 1b) (Pinto et al. 2022). Lastly, we extended our inferences to a pair of much less divergent systems within the infraorder Scincomorpha. Here—unlike mammals, birds, and lacertids—both species have homomorphic sex chromosome systems, or systems that are indistinguishable under a light microscope. In the Cape cliff lizard (*H. capensis*; XY), we see diagnostic differences

between their X and Y chromosomes that are reinforced by multiple rearrangements between the chromosomal pair when aligned (Fig. 3). And in the Christmas Island skink (*C. egeriae*; XY), we observed a pair of contigs deviating from autosomal background on a region corresponding to the known sex chromosome linkage group in chicken (Gg1) (Fig. 4). Embedded within these contigs is a small region that aligned poorly, the sex-limited region, corresponding to published expectations for the system in skinks (Kostmann et al. 2021). In passing, we note that within the publicly available contig assembly for the Christmas Island skink (GCA_030015325.1), there exists an equivalent contig to the chrX contig presented here, JAREYC01000008.1, while the putative sex-limited region of chrY is split off as its own unique contig, JAREYC01000009.1. Thus, both X/Z and Y/W gametologous regions possess detectably unique densities of hap-mers—not seen in their autosomal counterparts—in both highly degenerated heteromorphic systems that are readily diagnosable under a light microscope (eg mammals, birds,

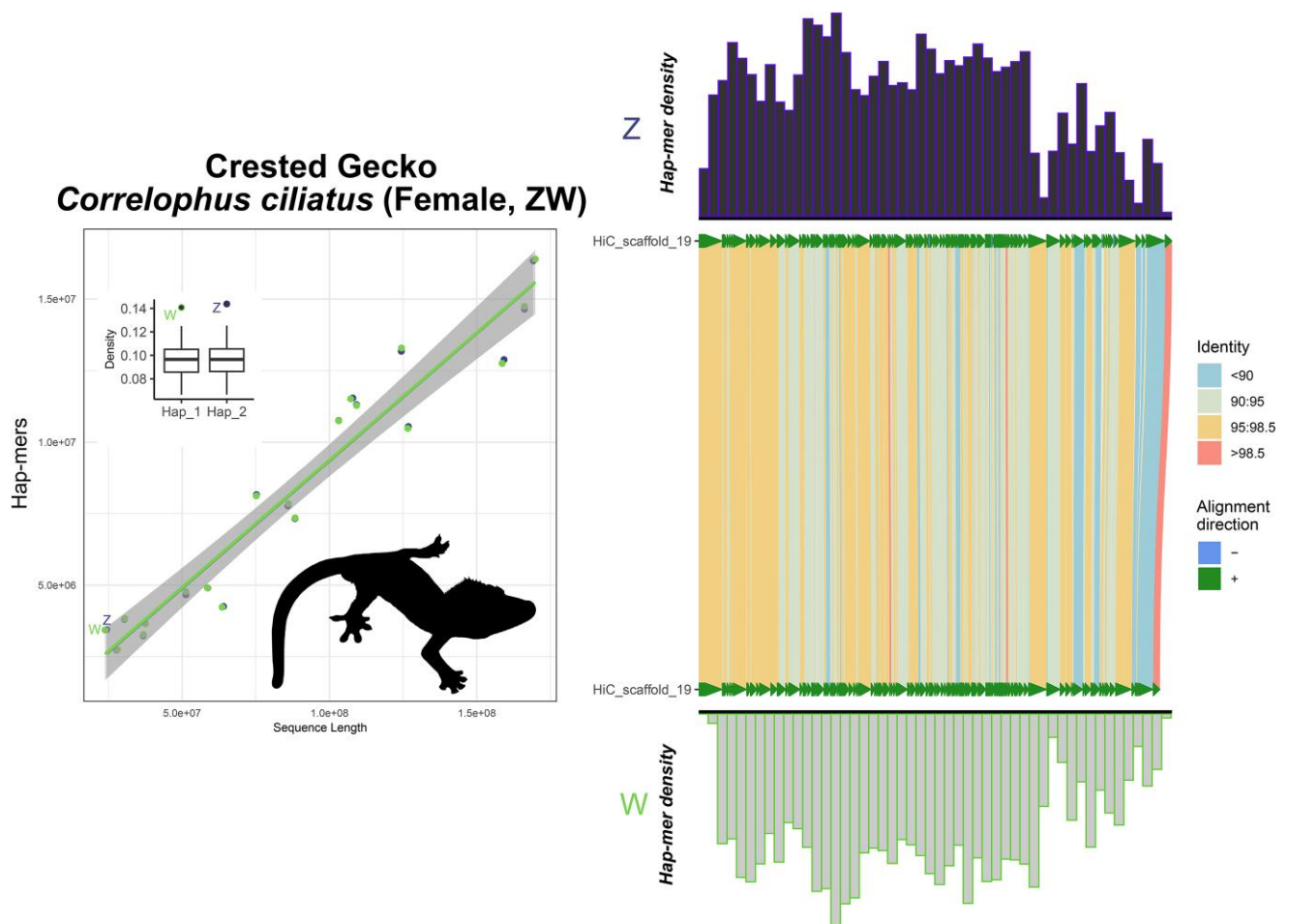


Figure 5 Total evidence plot for crested gecko, *Correlophus ciliatus*, under the SCINKD framework. The left panel consists of SCINKD “results” output ~ sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1 Mb windows across the chromosome. A single pair of haplotypes deviates from the autosomal background. The signal itself originates from LG19 across most of their length, a chromosomal pair that possesses generally low sequence identity across its length with the exception of a small region at the distal tip, consistent with a pseudo autosomal region (PAR). This evidence complements and is confirmed by previous work showing that LG19 is the ZW linkage group in the crested gecko and annotated the Z and W chromosomes using sex-specific RADtags (Gamble et al. 2015; Keating 2022).

and lacertids; Fig. 2), and those that are not, ie “homomorphic” systems (eg *Cryptoblepharus* and *Hemicordylus*; Figs. 3 and 4) (Charlesworth 2021).

Conceptual framework overview

Deeper interpretation of exemplar taxa in Fig. 2 among others—also supported by myriad embargoed datasets—led to a refined conceptual framework described in Fig. 1. Briefly, in regions of restricted recombination, degeneration of the sex-limited chromosome is expected due to reduced efficiency of germline DNA repair and driven evolutionarily by both selection and drift (Bergero and Charlesworth 2009). The absence of homologous recombination in the sex-limited region(s) of the sex chromosomes causes genetic divergence between gametologs (X and Y or Z and W), and this process of degeneration can happen over relatively short evolutionary timeframes. Although difficult to validate in the absence of

additional data (ie genomic information from multiple females and males), these differences often occur in excess relative to the autosomal background (Figs. S5 and S6). In heteromorphic systems, these differences are extreme (eg Fig. 2a, c, and e; Fig. S5) and can typically also be supported by read depth-related measurements. However, most vertebrate species possess homomorphic sex chromosomes (Hillis and Green 1990; Ezaz et al. 2009; Schultheis et al. 2009), but these systems still possess differences in excess relative to an autosomal background (Fig. S6). Further, in certain taxa with ancient sex chromosome systems (ie hundreds of millions of years of reduced effective population size), the sex chromosomes in a homogametic individual (XX or ZZ) can also be identified by a dearth of hap-mers (Fig. 2b and d) (Wilson Sayres 2018). Under certain conditions, there may be no excess or dearth of hap-mers present (or a correlation between chromosome length and hap-mer quantity fails to appear) and in these cases SCINKD would not be effective. These conceptual points are presented in a graphical framework in Fig. 1.

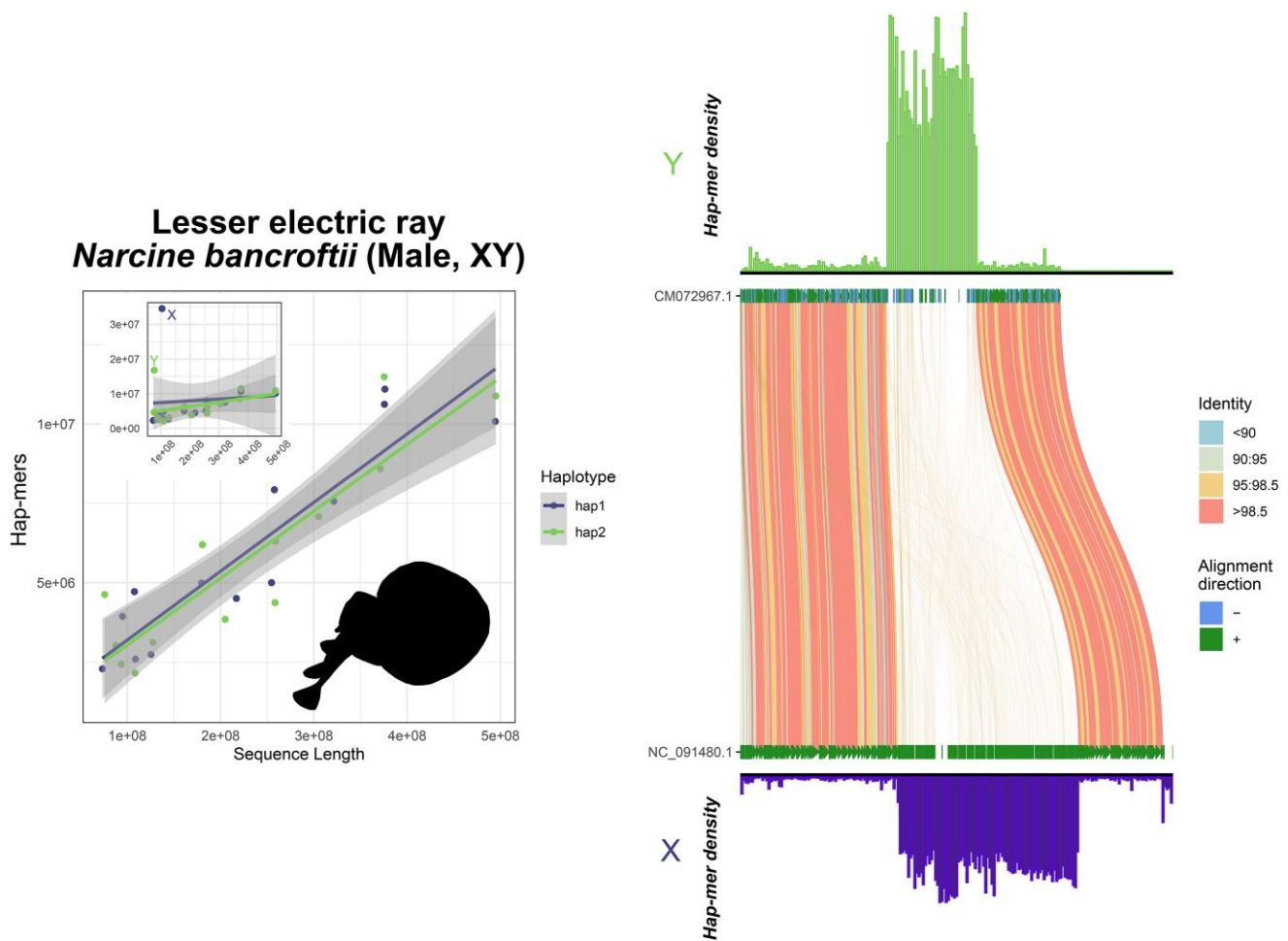


Figure 6 Total evidence plot for the lesser electric ray, *Narcine bancroftii*, under the SCINKD framework. The left panel consists of SCINKD “results” output \sim sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1 Mb windows across the chromosome. A single pair of haplotypes deviates from the autosomal background. The signal itself originates from a single region at the center of LG12, a region that aligns poorly between haplotypes and possesses an excess of unique kmers. The convergence of evidence on this region corroborates previous work showing this is the conserved XY linkage group in many elasmobranch species (Lee et al. 2025). Here, the X and Y are annotated based on homology with previously annotated X and Y in related species.

Divergent and previously unannotated test cases

When extrapolating our SCINKD conceptual framework across new data, we targeted a broad sample of numerous independent sex chromosome systems with varying levels of empirical support (Figs. 3–10).

1. In the crested gecko (*C. ciliatus*), our results support the previously identified ZW system using hap-mer densities, a system supported by RADseq data of multiple male and female samples (Fig. 5) (Gamble et al. 2015; Keating 2022).
2. In the lesser electric ray (*Narcine bancroftii*), we expanded further across the phylogeny to observe strong support for the presumed ancestral XY linkage group on LG12. This species was assumed to be XX/XY but due to ambiguity from other methods, no sex chromosomes have been previously annotated (Fig. 6) (Lee et al. 2025).

3. In the San Diegan legless lizard (*A. stebbinsi*), we see strong support for patterns consistent with a ZW system on LG7 (Fig. 7). This is consistent with other data suggesting that anguimorphs possessed an ancestral ZZ/ZW sex chromosome system with multiple transitions to novel ZW linkage groups (Rovatsos et al. 2019; Pinto et al. 2024).
4. In the Christmas Island gecko (*L. listeri*), we see low support for patterns consistent with an XY system on LG18 (Fig. 8). Relatively little data are available across this extraordinarily diverse family, Gekkonidae, to assess the likelihood or conservation of a putative XX/XY system in this gecko species. However, it is recently been hypothesized that *Lepidodactylus* possess a ZZ/ZW system, suggesting a misassembly on LG18 is likely the cause of this signal (Volobouev and Paster 1988; Gamble 2010).

Neither latter species, *A. stebbinsi* and *L. listeri*, has publicly available resequencing data from males and females to further validate these hypotheses. However, in cases like these

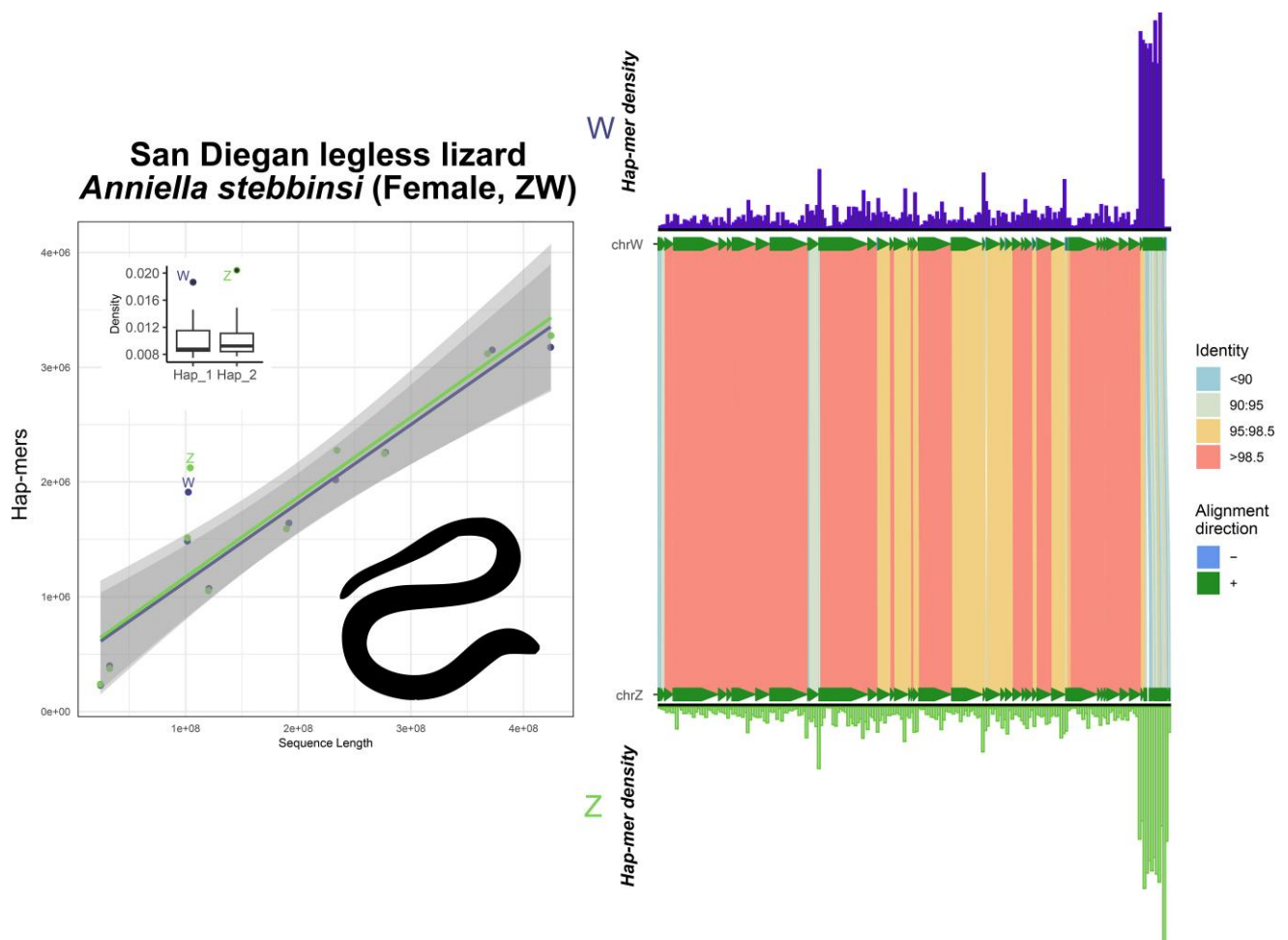


Figure 7 Total evidence plot for the San Diegan legless lizard, *Anniella stebbinsi*, under the SCINKD framework. The left panel consists of SCINKD “results” output ~ sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1 Mb windows across the chromosome. A single pair of haplotypes deviates from the autosomal background. The signal itself originates from a single region at the distal end of LG7, a region that possesses reduced sequence identity relative to the rest of the pairwise alignment. The convergence of evidence on a single region provides strong support for a hypothesis of a ZZ/ZW system in this anguimorph lizard. Here, the Z and W were hypothesized due to sequence length and additionally supported by alignment to the corresponding linkage group (NC_086184.1) in the Southern alligator lizard, *Elgaria multicarinata*, assembly (GCF_023053635.1), where average sequence identity was slightly higher on the putative Z-specific region (81%) than on the W-specific region (80%), which would be expected due to higher rates of sequence divergence on chrW.

where additional data are unavailable or prohibitive, our results suggest that SCINKD is more robust than a read depth-related approach in supporting candidate sex-linked regions in recently evolved sex chromosome systems (Pinto et al. 2022).

Next, to examine SCINKD robustness using a more diverse taxon, we tested SCINKD with an attempt to independently confirm the sex chromosome linkage group in a New Caledonia shrub species, *Amborella trichopoda*, with known ZZ/ZW sex chromosomes (Carey et al. 2024). In this shrub, we observe strong deviations from our vertebrate-focused model of correlations between hap-mers and chromosome length (Fig. 9). However, the ZW chromosomal pair still possess the most hap-mers. The lack of correlation between hap-mers and chromosome length can likely be attributed to the distinct mechanisms of recombination present in plants (Gaut et al. 2007), which is thought to drive much of the previously described correlations of chromosome

length in vertebrates (Pinto et al. 2023b). Regardless, since hap-mer densities appear to be disproportionate on the sex chromosomes in *Amborella*, there may be potential for developing a more broad SCINKD conceptual model that accounts for alternative biology present in nonvertebrate lineages (Fig. 9).

Lastly, we developed an opportunistic test case for a difficult assembly with a known XX/XY sex chromosome system in the Neotropical leaf-litter gecko, *S. notatus* (Pinto et al. 2022). In this experiment, the *S. notatus* sample contained technical artifacts from a combination of (i) low HiFi sequencing yield (~20× coverage), (ii) shorter-than-optimal read lengths (mean read length = ~13.5 kb), and (iii) a dearth of heterozygosity (0.35%) perhaps due to its hypothesized low effective population size (Clements et al. 2021)—which is even below the rate previously documented to cause phasing issues in other taxa propagated in the pet trade, eg 0.56% in *E. macularius* (Pinto et al. 2023a). Under these conditions, we see our first instance of a

Christmas island gecko *Lepidodactylus listeri* (Male, unknown)

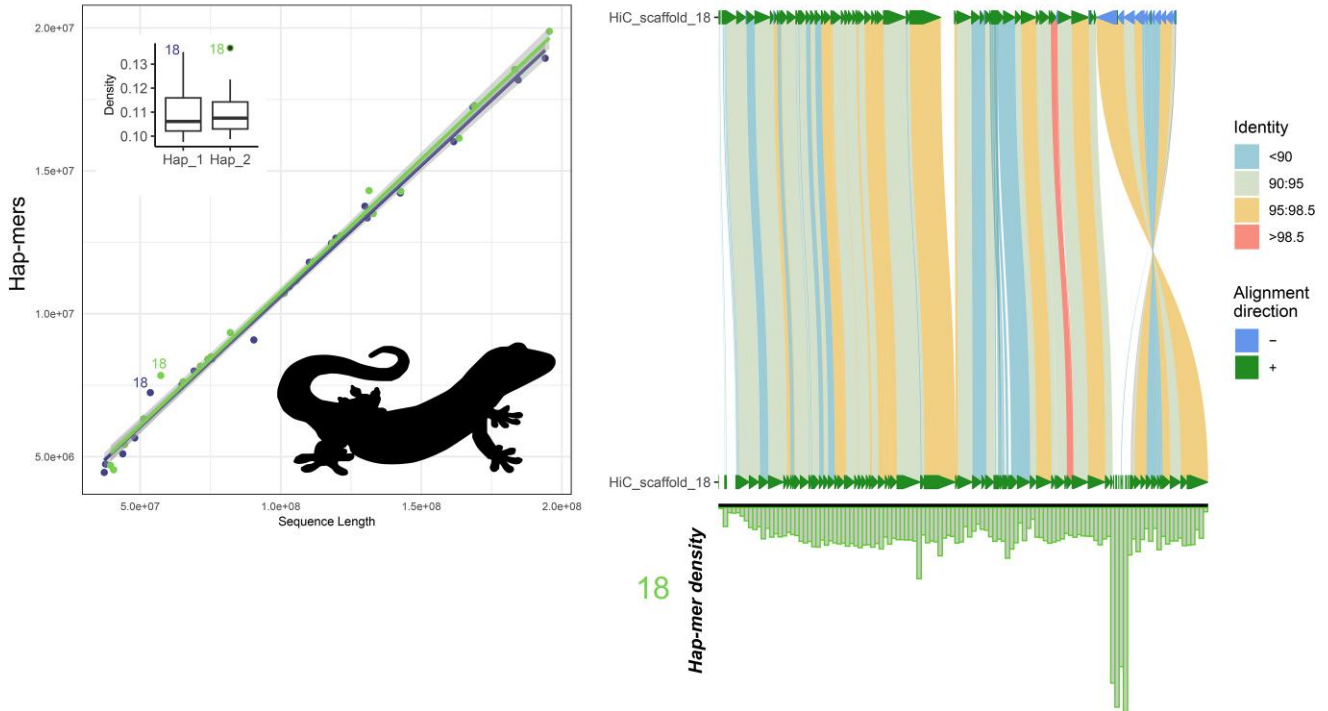


Figure 8 Total evidence plot for the Christmas Island gecko, *Lepidodactylus listeri*, under the SCINKD framework. The left panel consists of SCINKD “results” output ~ sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1 Mb windows across the chromosome. A single pair of haplotypes deviates from the autosomal background. The signal itself originates from two poorly aligned regions between the haplotype pairs. The convergence upon a linkage group of interest, but not focused on a specific region, provides relatively low support for a hypothesis that LG18 is the XY linkage group in this gecko species, but could warrant further investigation using gene annotations and data from additional individuals. Low confidence of support for this putative XY system pre-empts any attempt to assign haplotypes as putative X/Y chromosomes.

vertebrate deviating from expectations (Fig. 10). Here, chr1 is the known XY linkage group (Pinto et al. 2022). However, low coverage and technical artifacts prevent the haplotypes from conforming to the expected correlation between sequence length to hap-mer abundance and preempt SCINKD from isolating them. Indeed, because the SCINKD framework is a biological prediction, the diploid reference genome data representing that biology may possess inherent biases that can inhibit genome assembly and/or phasing on the sex chromosomes (Rhie et al. 2020; Cheng et al. 2022; Pinto et al. 2023a; Antipov et al. 2025). Thus, because there is a direct effect of sequencing depth on per-base quality and completeness of any given genome assembly, insufficient coverage may be a recurrent issue in many species.

What does SCINKD support mean?

SCINKD was specifically designed to identify homomorphic sex chromosome systems. Although by extension it works well in

heteromorphic systems, there are other robust methods developed to identify heteromorphic sex chromosomes, including cytogenetics and read depth analysis (<https://github.com/makovalab-psu/HalfDeep>). SCINKD provides a multidimensional breakdown of haplotype uniqueness that helps inform the likelihood of sex-limited regions in the genome. At a global scale (dot plots and box plots), we develop a robust framework to investigate statistical outliers (eg Fig. 7). After a candidate(s) is identified, one can assess hap-mer density on a finer scale using mirror plots (eg Figs. S3 to S5). Lastly, corroborating these layers with 1:1 chromosome alignments for candidate linkage groups (eg using minimap2 and SVbyEye) provides a link between chromosome coordinates and haplotype-specific anomalies (Fig. 7). Indeed, in *A. stebbinsi*, a homomorphic system we view as providing strong evidence under this framework, each of the aforementioned analyses converges on a single set of haplotypes where hap-mer density increases exponentially and sequence identity between haplotypes drops sharply (Fig. 7). These factors taken together provide evidence to hypothesize a ZZ/ZW system in *A. stebbinsi*.

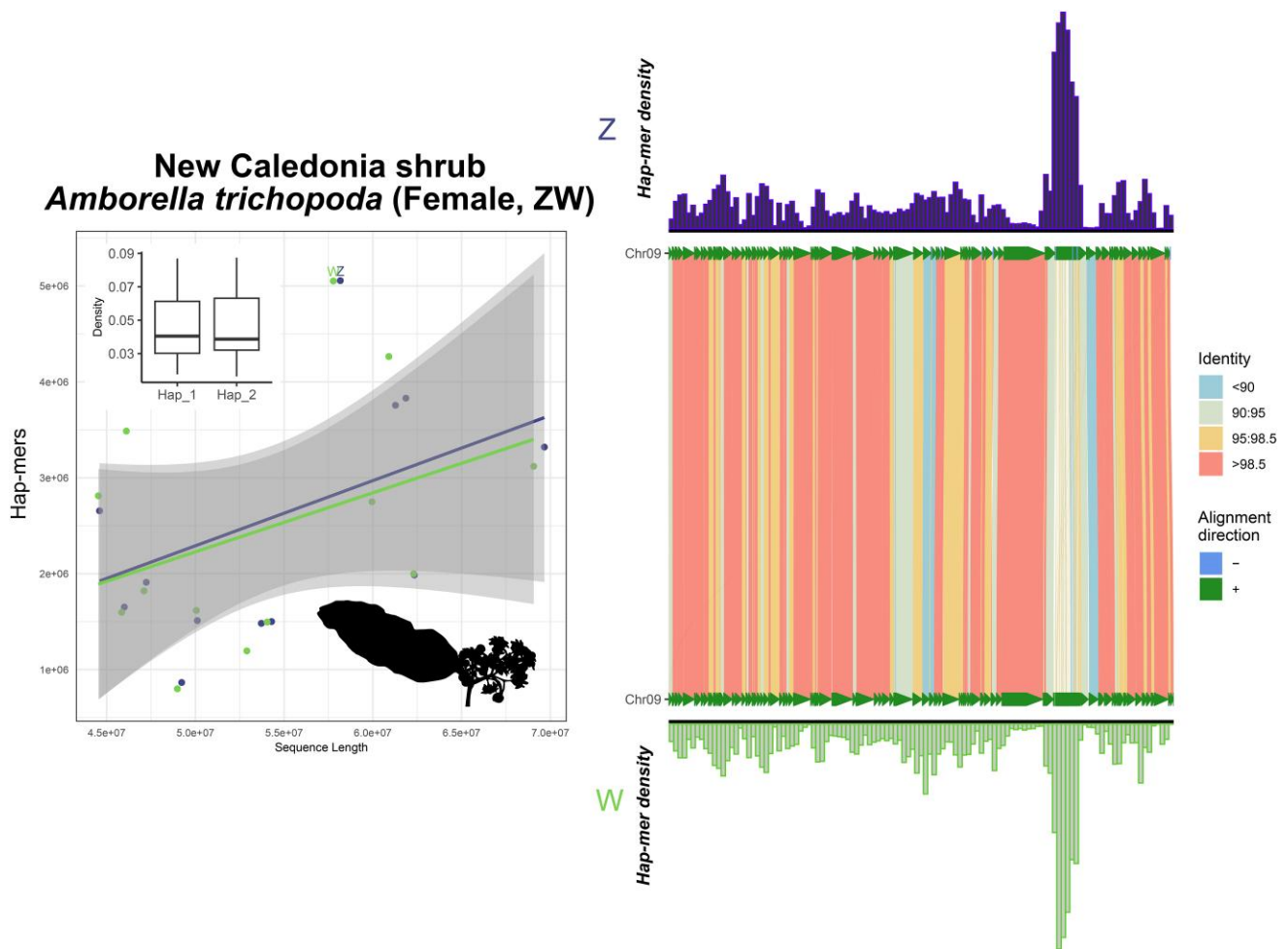


Figure 9 Total evidence plot for the New Caledonia shrub, *Amborella trichopoda*, under the SCINKD framework. The left panel consists of SCINKD “results” output ~ sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1Mb windows across the chromosome. A single pair of haplotypes deviates from the autosomal background. The signal itself originates from a single region at the distal end of LG7, a region that possesses reduced sequence identity relative to the rest of the pairwise alignment. The haplotype density and pairwise alignment in the previously identified sex-limited region is concordant with previous work showing these are the Z and W chromosomes in this species of shrub (Carey et al. 2024). However, no robust statistical support is observed under the SCINKD framework, suggesting plants do not conform to our vertebrate-centric linear predictions of haplotype divergence.

Caveats and limitations

We hypothesize that phasing issues may be the primary cause for excessive noise observed under the proposed framework. We show here that sequencing depth itself, above a reasonable threshold ($>20\times$), does not greatly affect sex chromosome signals in the high-quality sequencing data for the putatively homomorphic ZW system in *A. stebbinsi* (Fig. S2). However, misassemblies as small as 1 Mb may begin to obscure sex chromosome signals (Fig. S1). Thus, as long as a sequencing depth and heterozygosity are great enough to accurately phase the genome, SCINKD may be an exceedingly useful tool for sex chromosome identification. We show that per-haplotype sequencing depth needs to exceed the standard of $10\times$ coverage using PacBio HiFi reads to unobscure sex chromosome signals in our *A. stebbinsi* example (Fig. S2). However, greater sequencing depth reduces errors and can help prevent misphasing and collapses, which can obscure

otherwise clear signals from the sex chromosomes (Fig. S1). It remains untested how equivalent ONT data performs in this context. A more robust sensitivity analysis in the future is warranted when a broad taxonomic sampling of high-coverage ONT data is publicly available. Preliminary data shared with us across diverse taxa suggest that $30\times$ ONT coverage may not be sufficient to prevent the aforementioned technical artifacts that interfere with the underlying assumptions of the SCINKD framework (A. Barley and L. Gray personal communication), although $100\times$ coverage is more than sufficient to identify the heteromorphic sex chromosome system in the Gila monster (unpublished data). In sum, to help address technical artifacts that interfere with haplotype resolution across the genome, but especially the sex chromosomes where the sex-limited region is assembled largely independently, we believe that recommending coverage at $20\times$ to $\sim 30\times$ per-haplotype ($40\times$ to $\sim 60\times$ total coverage) is justifiable to assist in identification of sex chromosomes in many diverse taxa. This recommendation coincides

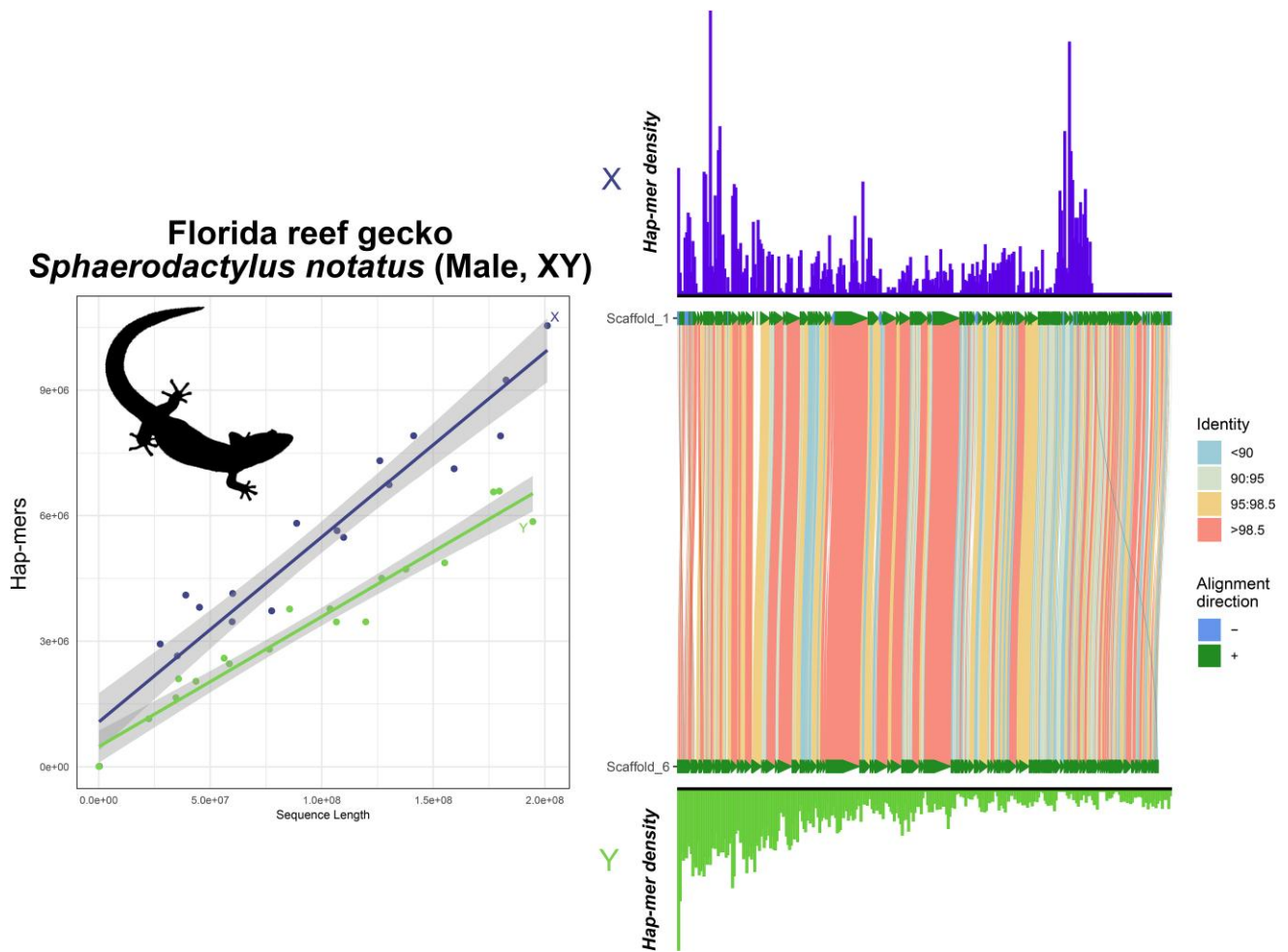


Figure 10 Total evidence plot for the Florida reef gecko, *Sphaerodactylus notatus*, under the SCINKD framework. The left panel consists of SCINKD “results” output ~ sequence length (Li et al. 2009) visualized with ggplot2 (Wickham 2016). The right panel corresponds to a minimap2 alignment (Li 2018) of putative sex chromosome haplotypes visualized using SVbyEye (Porubsky et al. 2025) overlaid with hap-mer densities in 1 Mb windows across the chromosome. We see no explicit evidence of the XX/XY system in this species under the SCINKD framework. A lack of convergence upon a common slope suggests phasing issues with the genome assembly. The sex chromosome linkage group itself, identified by previous work (Pinto et al. 2022), also shows no evidence of sex chromosome patterns in the sex-limited region suggesting an excess of noise, likely caused by a combination of low sequencing coverage (~20x) and a dearth of heterozygosity. However, we were able to annotate the X and Y chromosomes using previously published sex-specific RADtags (Pinto et al. 2022).

with recent increases in sequencing outputs and deviates from the current recommended standards of 30x total coverage for a typical vertebrate assembly (Pinto et al. 2023b).

The data do not yet currently exist to characterize every possible outcome across divergent taxa, technical capacity, or sex chromosome system. Observations presented here also suggest that the Y/W typically falls lower on both axes of the dot plot, irrespective of chromosome length. This is presumably due to (i) the sex-limited chromosome (W or Y) being typically shorter than its gametologous counterpart, and (ii) accumulation of repetitive DNA and loss of gene content, both of which can reduce the numbers of unique hap-mers on the Y/W relative to their X/Z homolog, and relative to their own length. However, as is often the case in biology, this is not always true and should not be used as the only metric for sex chromosome assignment. Another notable observation here is that no two sex chromosome alignments share any notable feature that would independently indicate assignment as a sex chromosome (Figs. 2 to 10), which is why using a pairwise

alignment alone to identify candidate loci across diverse systems is not sufficient. Thus, given the theoretical framework presented here (Fig. 1) and the ability to view multivariate outputs from the SCINKD workflow (ie dotplots, density boxplots, and mirror plots), combined with pairwise alignments, we provide researchers the ability to make strong hypotheses regarding sex chromosome assignments across diverse taxa from a single, diploid genome. We emphasize the importance of an empirical approach (an understanding of the biology of the study system as well as sex chromosome evolution) when applying this framework.

How should these sex chromosomes be annotated?

It is worth mentioning a few important observations of the data presented here. No individual analysis (eg HiC data, read depth,

haplotype alignment, or SCINKD results) can be taken alone as conclusive evidence that a pair of haplotypes make up a novel sex chromosome system. Indeed, when applying methods such as SCINKD at scale, we will continue to provide strong evidence for novel putative sex chromosomes where no other data are available for validation, such as in the San Diegan legless lizard (Fig. 7). We suggest a four-category system for sex chromosome annotation for any particular genome, and cluster taxa from this study into them as examples. We describe two broad scenarios, differentiated by levels of available evidence, but always recommend cross-referencing using data from additional individuals when available.

1. In species with strong evidence of sex chromosomes, the sex chromosomes should be annotated, (A) the “gold standard”—those validated by external data types (such as resequencing data, RADseq, or a high-quality assembly of both sexes, etc.)—eg mammals/birds/lacertids (Fig. 2), cape cliff lizard (Fig. 3) (Leitão et al. 2023), crested gecko (Fig. 5) (Gamble et al. 2015), New Caledonia shrub (Fig. 9) (Carey et al. 2024), and Florida reef gecko (Fig. 10) (Pinto et al. 2022), or (B) corresponding evidence from SCINKD and another source(s) (such as phylogenetic history and/or read depth from the assembly input data)—eg Christmas Island skink (Fig. 4), lesser electric ray (Fig. 6), and San Diegan legless lizard (Fig. 7).
2. In species with moderate-to-low evidence of sex chromosome linkage information, the sex chromosomes should “not” be annotated and await additional data (such as resequencing data, a high-quality assembly of the opposite sex, RADseq, etc.), such as (C) those assemblies with a statistically predicted candidate linkage group with no candidate sex-determining region or phylogenetic information—eg the Christmas Island gecko (Fig. 8), or (D) those with no clear evidence of a sex chromosome linkage group in the assembly—eg the ZZ Nicobar pigeon (Fig. 2d), ZZ Cretan wall lizard (Fig. 2f), or the TSD leopard gecko (Fig. S4).

Here, we present myriad examples from typical vertebrate (and a plant) sex chromosome systems with varying levels of success. However, sex determination itself is a complicated process and many species do not even possess sex chromosomes. For example, given that hymenopterans possess complementary sex determination, we can confidently state that well-over 20% of total extant animal species do not possess sex chromosomes (Whiting 1943; Heimpel and de Boer 2008). Outside of Metazoa, ~5% to ~10% of angiosperms use genetic sex determination (Charlesworth and Harkess 2024), but to our knowledge *A. trichopoda* is the only plant species to date with assembled sex chromosomes in a haplotype-resolved assembly. Beyond that, other aspects of life history vary wildly across taxa, from TSD to polygenic sex determination to species utilizing a sex-limited mitochondrial region as a locus of large effect (Smith et al. 2023; Kocher et al. 2024). In any of these instances, it is possible that SCINKD may assist in genome curation but would be improperly utilized to hypothesize the existence of sex chromosomes in these types of diverse taxa. Thus, since the null predictions used in SCINKD are biological with technical considerations applied to interpretation, it is important to understand the biology of each system prior to generating any hypotheses using SCINKD.

Conclusion

The utility of SCINKD in the age of biodiversity genomics is apparent and has already been deployed to assist with sex chromosome curation across the most recent genome assemblies from the Vertebrate Genomes Project (VGP) (Rhie et al. 2021). However, there also remain vertebrate lineages that remain underrepresented by the VGP, such as squamates, the most speciose discrete clade of tetrapods (Rhie et al. 2021; Pinto et al. 2023b). As these lesser-known taxa across the tree of life begin acquiring reference genomes, there will continue to be a lack of population-level genomic data traditionally used to develop sex chromosome annotations. We view the SCINKD framework as a significant step forward toward addressing the ongoing challenges in sex chromosome assembly and annotation at a scale to meet the rapidly moving field of genomics.

Acknowledgments

The authors would like to acknowledge M. Kirkpatrick and J.C. Havird for additional funding and laboratory space for the *S. notatus* HiC data and comments on the manuscript; B. Pickett for assistance in navigating meryl syntax between SCINKD versions 1 and 2; S. Plaisier for discussion on the nascent concept of SCINKD; and L. Gray and A. Barley for sharing unpublished ONT genomes and data for additional sensitivity testing. For novel data presented here, animal husbandry and tissue sampling were conducted under Marquette University IACUC protocol AR-279. The authors acknowledge Research Computing at Arizona State University for providing high performance computing resources that have contributed to the research results reported within this paper. Figure silhouettes were provided by PhyloPic (<https://www.phylopic.org/>) and are provided in the public domain for reuse.

Author contributions

B.J.P. conceptualized the project, acquired funding, developed the tool, conducted genome sequencing and assembly experiments, and drafted the manuscript; S.M.G. ran vertebrate-wide analyses and generated data figures; S.E.K. generated RADseq library; C.H.S. acquired funding and assisted with HiC experiment; T.G. provided materials and laboratory space; S.V.N. acquired funding, feedback, and the *Sphaerodactylus* sample; and M.A.W. acquired funding, supervised the project, and provided regular feedback.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

This work was funded piecemeal by the University of Texas at Austin Office of the Vice President (OVPR) Special Research Grant to B.J.P. and M. Kirkpatrick; National Science Foundation (USA) DEB1657662 and IOS2151318 to T.G.; the University of

Texas at Austin Stengl-Wyer Endowment to C.H.S.; National Institute of Health (USA) R35GM142836 to J.C. Havird; and National Institute of Health National Institute of General Medical Sciences (USA) MIRA R35GM124827 to M.A.W. This work was also supported, in part, by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health.

Conflict of interest

The authors declare no conflict of interest.

Data availability

Both SCINKD code and user tutorial are available on GitHub (<https://github.com/DrPintoThe2nd/SCINKD>). Raw data generated in this study is available on SRA (SAMN48008689-90) and genome reassemblies on Figshare (<https://doi.org/10.6084/m9.figshare.29344685>) and, where applicable, also to GenBank.

References

- Alonge M *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 2022;23:258. <https://doi.org/10.1186/s13059-022-02823-7>.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Antipov D *et al.* Verkko2 integrates proximity-ligation data with long-read De Bruijn graphs for efficient telomere-to-telomere genome assembly, phasing, and scaffolding. *Genome Res.* 2025;35:1583–1594. <https://www.genome.org/cgi/doi/10.1101/gr.280383.124>.
- Behrens KA, Koblmüller S, Kocher TD. Diversity of sex chromosomes in vertebrates: six novel sex chromosomes in basal haplochromines (Teleostei: Cichlidae). *Genome Biol Evol.* 2024;16:evae152. <https://doi.org/10.1093/gbe/evae152>.
- Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol.* 2009;24:94–102. <https://doi.org/10.1016/j.tree.2008.09.010>.
- Card DC, Jennings WB, Edwards SV. Genome evolution and the future of phylogenomics of non-avian reptiles. *Animals (Basel).* 2023;13:471. <https://doi.org/10.3390/ani13030471>.
- Carey SB *et al.* ZW sex chromosome structure in Amborella trichopoda. *Nat Plants.* 2024;10:1944–1954. <https://doi.org/10.1038/s41477-024-01858-x>.
- Carvalho AB, Clark AG. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res.* 2013;23:1894–1907. <https://doi.org/10.1101/gr.156034.113>.
- Charlesworth D. The timing of genetic degeneration of sex chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 2021;376:20200093. <https://doi.org/10.1098/rstb.2020.0093>.
- Charlesworth D, Harkess A. Why should we study plant sex chromosomes? *Plant Cell.* 2024;36:1242–1256. <https://doi.org/10.1093/plcell/koad278>.
- Cheng H *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol.* 2022;40:1332–1335. <https://doi.org/10.1038/s41587-022-01261-x>.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Clery JG *et al.* 2015. Comparing variant call files for performance benchmarking of next-generation sequencing variant calling pipelines [preprint]. bioRxiv 023754. <https://doi.org/10.1101/023754>.
- Clements SL, Powell EA, Mothes CC, Searcy CA. Assessing the conservation risk of *Sphaerodactylus notatus*, the U.S. herpetofaunal species most vulnerable to sea level rise. *Biodivers Conserv.* 2021;30:107–121. <https://doi.org/10.1007/s10531-020-02080-9>.
- Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics.* 2017;18:341. <https://doi.org/10.1186/s12864-017-3723-5>.
- Dodge TO *et al.* Genomes of two extinct-in-the-wild reptiles from Christmas Island reveal distinct evolutionary histories and conservation insights. *Mol Ecol Resour.* 2025;25:e13780. <https://doi.org/10.1111/1755-0998.13780>.
- Durand NC *et al.* Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Syst.* 2016;3:95–98. <https://doi.org/10.1016/j.cels.2016.07.002>.
- Ezaz T, Sarre SD, O’Meally D, Graves JA, Georges A. Sex chromosome evolution in lizards: independent origins and rapid transitions. *Cytogenet Genome Res.* 2009;127:249–260. <https://doi.org/10.1159/000300507>.
- Feron R *et al.* RADSex: a computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol Ecol Resour.* 2021;21:1715–1731. <https://doi.org/10.1111/1755-0998.13360>.
- Gable SM *et al.* The state of squamate genomics: past, present, and future of genome research in the most speciose terrestrial vertebrate order. *Genes (Basel).* 2023;14:1387. <https://doi.org/10.3390/genes14071387>.
- Gamble T. A review of sex determining mechanisms in geckos (Gekkota: Squamata). *Sex Dev.* 2010;4:88–103. <https://doi.org/10.1159/000289578>.
- Gamble T *et al.* Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among Gecko sex-determining systems. *Mol Biol Evol.* 2015;32:1296–1309. <https://doi.org/10.1093/molbev/msv023>.
- Gamble T, Zarkower D. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol Ecol Resour.* 2014;14:902–913. <https://doi.org/10.1111/1755-0998.12237>.
- Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet.* 2007;8:77–84. <https://doi.org/10.1038/nrg1970>.
- Graves JAM. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet.* 2016;17:33–46. <https://doi.org/10.1038/nrg.2015.2>.
- Gumangan MA, Pan Z, Lozito TP. Chromosome-level genome assembly and annotation of the crested gecko, *Correlophus*

- ciliatus*, a lizard incapable of tail regeneration. *GigaByte*. 2024:2024:gigabyte140. <https://doi.org/10.46471/gigabyte.140>.
- Heimpel GE, de Boer JG. Sex determination in the hymenoptera. *Annu Rev Entomol*. 2008;53:209–230. <https://doi.org/10.1146/annurev.ento.53.103106.093441>.
- Hillis DM, Green DM. Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. *J Evol Biol*. 1990;3:49–64. <https://doi.org/10.1046/j.1420-9101.1990.3010049.x>.
- Jeffries D *et al*. The tree of sex consortium: a global initiative for studying the evolution of reproduction in eukaryotes. *J Evol Biol*. 2025;38::861–886. <https://doi.org/10.1093/jeb/voaf053>.
- Keating SE. Evolution of sex chromosomes in Geckos (Reptilia: Gekkota). [Dissertations]. Marquette University; 2022. https://epublications.marquette.edu/dissertations_mu/1562/
- Kocher TD, Meisel RP, Gamble T, Behrens KA, Gammerdinger WJ. Yes, polygenic sex determination is a thing!. *Trends in Genetics*. 2024;40:1001–1017. <https://doi.org/10.1016/j.tig.2024.10.003>.
- Kostmann A, Kratochvíl L, Rovatsos M. Poorly differentiated XX/X Y sex chromosomes are widely shared across skink radiation. *Proc Biol Sci*. 2021;288:20202139. <https://doi.org/10.1098/rspb.2020.2139>.
- Lee S-H *et al*. 2025. Insights into the evolution of ancient shark and ray sex chromosomes [preprint]. bioRxiv 637739. <https://doi.org/10.1101/2025.02.26.637739>.
- Leitão HG, Diedericks G, Broeckhoven C, Baeckens S, Svoldal H. Chromosome-level genome assembly of the cape cliff lizard (*Hemicordylus capensis*). *Genome Biol Evol*. 2023;15:evad001. <https://doi.org/10.1093/gbe/evad001>.
- Li H *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Linck EB, Cadena CD. A latitudinal gradient of reference genomes. *Mol Ecol*. 2024;34:e17551. <https://doi.org/10.1111/mec.17551>.
- Makova KD *et al*. The complete sequence and comparative analysis of ape sex chromosomes. *Nature*. 2024;630:401–411. <https://doi.org/10.1038/s41586-024-07473-2>.
- Marshall Graves JA. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu Rev Genet*. 2008;42:565–586. <https://doi.org/10.1146/annurev.genet.42.110807.091714>.
- Nielsen SV *et al*. Escaping the evolutionary trap? Sex chromosome turnover in basilisks and related lizards (Corytophanidae: Squamata). *Biol Lett*. 2019;15:20190498. <https://doi.org/10.1098/rsbl.2019.0498>.
- Nishimura O, Hara Y, Kuraku S. Gvolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*. 2017;33:3635–3637. <https://doi.org/10.1093/bioinformatics/btx445>.
- Palmer DH, Rogers TF, Dean R, Wright AE. How to identify sex chromosomes and their turnover. *Mol Ecol*. 2019;28:4709–4724. <https://doi.org/10.1111/mec.15245>.
- Pinto BJ *et al*. A chromosome-level genome assembly of the parasitoid wasp, *Cotesia glomerata* (Hymenoptera: Braconidae). *J Hered*. 2021;112:558–564. <https://doi.org/10.1093/jhered/esab032>.
- Pinto BJ *et al*. Chromosome-level genome assembly reveals dynamic sex chromosomes in neotropical leaf-litter geckos (Sphaerodactylidae: *Sphaerodactylus*). *J Hered*. 2022;113:272–287. <https://doi.org/10.1093/jhered/esac016>.
- Pinto BJ *et al*. The revised reference genome of the leopard gecko (*Eublepharis macularius*) provides insight into the considerations of genome phasing and assembly. *J Hered*. 2023a;114:513–520. <https://doi.org/10.1093/jhered/esad016>.
- Pinto BJ *et al*. It's a trap?! Escape from an ancient, ancestral sex chromosome system and implication of *Foxl2* as the putative primary sex-determining gene in a lizard (Anguimorpha; Shinisauridae). *Evolution*. 2024;78:355–363. <https://doi.org/10.1093/evolut/qp205>.
- Pinto BJ, Gamble T, Smith CH, Wilson MA. A lizard is never late: Squamate genomics as a recent catalyst for understanding sex chromosome and microchromosome evolution. *J Hered*. 2023b;114:445–458. <https://doi.org/10.1093/jhered/esad023>.
- Pinto BJ, O'Connor B, Schatz MC, Zarate S, Wilson MA. Concerning the eXclusion in human genomics: the choice of sex chromosome representation in the human genome drastically affects the number of identified variants. *G3 (Bethesda)*. 2023c;13:jkad169. <https://doi.org/10.1093/g3journal/jkad169>.
- Porubsky D *et al*. SVbyEye: A visual tool to characterize structural variation among whole-genome assemblies. *Bioinformatics*. 2025;41:btaf332. <https://doi.org/10.1093/bioinformatics/btaf332>.
- Raina VB, Jessop A, Greene EC. Biochemical mechanisms of genetic recombination and DNA repair. *Annu Rev Biochem*. 2025;94:161–193. <https://doi.org/10.1146/annurev-biochem-083024-113931>.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432. <https://doi.org/10.1038/s41467-020-14998-3>.
- Rhie A *et al*. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;592:737–746. <https://doi.org/10.1038/s41586-021-03451-0>.
- Rhie A *et al*. The complete sequence of a human Y chromosome. *Nature*. 2023;621:344–354. <https://doi.org/10.1038/s41586-023-06457-y>.
- Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21:245. <https://doi.org/10.1186/s13059-020-02134-9>.
- Rovatsos M *et al*. Conservation of sex chromosomes in lacertid lizards. *Mol Ecol*. 2016;25:3120–3126. <https://doi.org/10.1111/mec.13635>.
- Rovatsos M, Reháč I, Velenský P, Kratochvíl L. Shared ancient sex chromosomes in varanids, beaded lizards, and alligator lizards. *Mol Biol Evol*. 2019;36:1113–1120. <https://doi.org/10.1093/molbev/msz024>.
- Sansam CL, Pezza RJ. Connecting by breaking and repairing: mechanisms of DNA strand exchange in meiotic recombination. *FEBS J*. 2015;282:2444–2457. <https://doi.org/10.1111/febs.13317>.
- Schultheis C, Böhne A, Schartl M, Volff JN, Galiana-Arnoux D. Sex determination diversity and sex chromosome evolution in poeciliid fish. *Sex Dev*. 2009;3:68–77. <https://doi.org/10.1159/000223072>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation

- completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Smith CH *et al*. Mitonuclear sex determination? Empirical evidence from bivalves. *Mol Biol Evol*. 2023;40:msad240. <https://doi.org/10.1101/2023.07.05.547839>.
- Volobouev V, Pasteur G. Presumptive sex chromosomes of a unisexual homomorphic species of lizards, *Lepidodactylus lugubris*. *Heredity (Edinb)*. 1988;60:463–467. <https://doi.org/10.1038/hdy.1988.65>.
- Whiting PW. Multiple alleles in complementary sex determination of *Habrobracon*. *Genetics*. 1943;28:365–382. <https://doi.org/10.1093/genetics/28.5.365>.
- Wickham H. *Ggplot2: elegant graphics for data analysis*. Springer; 2016.
- Wilcoxon F. Individual comparisons by ranking methods. *Biom Bull*. 1945;1:80–83. <https://doi.org/10.2307/3001968>.
- Wilson Sayres MA. Genetic diversity on the sex chromosomes. *Genome Biol Evol*. 2018;10:1064–1078. <https://doi.org/10.1093/gbe/evy039>.