




Phased Assembly of Neo-Sex Chromosomes Reveals Extensive Y Degeneration and Rapid Genome Evolution in *Rumex hastatulus*

Bianca Sacchi ^{1,†*} Zoë Humphries ^{1,†*} Jana Kružlicová,^{2,3} Markéta Bodláková,² Cassandre Pyne,¹ Baharul I. Choudhury,^{1,4} Yunchen Gong,⁵ Václav Bačovský,² Roman Hobza,² Spencer C.H. Barrett,¹ and Stephen I. Wright ^{1,5,*}

¹Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

²Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic

³National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic

⁴Department of Biology, Queen's University, Kingston, Canada

⁵Centre for Analysis of Genome Evolution and Function, University of Toronto, Toronto, Canada

[†]These authors contributed equally.

***Corresponding authors:** E-mails: bianca.sacchi@mail.utoronto.ca; zoe.humphries@mail.utoronto.ca; stephen.wright@utoronto.ca.

Associate editor: Melissa Wilson

Abstract

Y chromosomes are thought to undergo progressive degeneration due to stepwise loss of recombination and subsequent reduction in selection efficiency. However, the timescales and evolutionary forces driving degeneration remain unclear. To investigate the evolution of sex chromosomes on multiple timescales, we generated a high-quality phased genome assembly of the massive older (<10 MYA) and neo (<200,000 yr) sex chromosomes in the XYY cytotype of the dioecious plant *Rumex hastatulus* and a hermaphroditic outgroup *Rumex salicifolius*. Our assemblies, supported by fluorescence in situ hybridization, confirmed that the neo-sex chromosomes were formed by two key events: an X-autosome fusion and a reciprocal translocation between the homologous autosome and the Y chromosome. The enormous sex-linked regions of the X (296 Mb) and two Y chromosomes (503 Mb) both evolved from large repeat-rich genomic regions with low recombination; however, the complete loss of recombination on the Y still led to over 30% gene loss and major rearrangements. In the older sex-linked region, there has been a significant increase in transposable element abundance, even into and near genes. In the neo-sex-linked regions, we observed evidence of extensive rearrangements without gene degeneration and loss. Overall, we inferred significant degeneration during the first 10 million years of Y chromosome evolution but not on very short timescales. Our results indicate that even when sex chromosomes emerge from repetitive regions of already-low recombination, the complete loss of recombination on the Y chromosome still leads to a substantial increase in repetitive element content and gene degeneration.

Key words: sex chromosomes, plants, genomics, transposable elements.

Introduction

One of the most striking patterns in genome evolution is the parallel degeneration of the nonrecombining chromosomes of the heterogametic sex (Y and W chromosomes, hereafter “Y”). Sex chromosomes have originated repeatedly across eukaryotes, and while far from universal, signatures of large-scale accumulation of deleterious mutations, the accumulation of repetitive elements, and the loss of gene function represent parallel evolutionary outcomes on the nonrecombining Y chromosome (Bachtrog 2013; Abbott et al. 2017). Although the extent of degeneration

varies greatly among species, many ancient Y chromosomes have lost nearly all their ancestral genes, with evidence of gene retention and sometimes expansion for genes important in reproductive function (Peichel et al. 2020; Subrini and Turner 2021) and meiotic drive (Bachtrog 2020). Despite the widespread recurrent nature of degeneration, our understanding of the timescales over which this occurs, and the evolutionary forces driving Y degeneration remains incomplete.

Several nonmutually exclusive evolutionary processes are thought to contribute to Y degeneration. First, the cessation of recombination causes widespread Hill–Robertson

Received: October 06, 2023. **Revised:** January 31, 2024. **Accepted:** April 05, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

interference between selected sites, weakening the efficacy of natural selection and driving the accumulation of slightly deleterious mutations (Rice 1987; Charlesworth et al. 2005). The loss of recombination can also cause a weakening of selection against transposable elements (TEs), both due to Hill–Robertson interference and a reduction in rates of ectopic recombination (Kent et al. 2017). Second, *cis*-regulatory divergence between the X and Y chromosome can drive loss of gene expression on the Y, enabling a positive feedback loop of expression loss and deleterious mutation accumulation on the nonrecombining sex chromosome that can occur even when Hill–Robertson interference effects are weak or absent (Lenormand et al. 2020). Positive selection for gene silencing or loss may also occur on the Y chromosome in cases where the retention of the Y gametolog confers a reduction in fitness, for example due to faster rates of adaptation on the X chromosome because of its higher effective population size (Orr and Kim 1998; Crowson et al. 2017) and/or the toxic effects of TE activity near genes on the Y (Wei et al. 2020; Muyle et al. 2022). Distinguishing the relative importance of these forces can be challenging, but an improved understanding of the earliest stages of Y degeneration can provide important insights.

The flowering plant *Rumex hastatulus* (Polygonaceae) represents an excellent model system for investigating the timescales and processes driving recombination suppression and Y degeneration. The species has two distinct heteromorphic sex chromosome cytotypes across its geographic range; males to the west of the Mississippi river have one X and one Y chromosome (XY cytotype). Based on our most recent phylogenomic analysis, this sex chromosome system is estimated to have arisen approximately 5 to 10 MYA (Hibbins et al. 2023). In contrast, males to the east of the Mississippi have an additional Y chromosome (XYY cytotype), the result of at least one reciprocal translocation event involving the X chromosome and one of the ancestral autosomes (Smith 1964; Kasjaniuk et al. 2019; Rifkin et al. 2021) approximately 180,000 yr ago (Beaudry et al. 2020). Our previous work suggested that the sex-linked regions in this species arose from large tracts of low recombination, particularly in male meiosis, which may have facilitated the evolution of large heteromorphic sex chromosomes (Rifkin et al. 2021, 2022). This includes the neo-sex-linked region, which arose from a region of reduced recombination on an ancestral autosome (Rifkin et al. 2021). This system creates an interesting opportunity to study the evolution of sex chromosome regions arising at different (but both young) timescales within the same genetic background.

To better understand the early stages of sex chromosome evolution and Y degeneration, we present a high-quality, fully phased assembly of the male genome of the XYY cytotype of *R. hastatulus* with highly contiguous assemblies of both Y chromosomes and the fused X chromosome. We characterize the patterns of chromosomal rearrangements, gene loss, and the repetitive DNA accumulation associated with sex chromosome evolution over multiple timescales in this genome. We also sequence

and assemble a hermaphroditic species in the genus, *Rumex salicifolius*, to infer changes in gene order and gene presence/absence evolution on the X and Y chromosomes.

Results and Discussion

Genome Assemblies

Our phased male genome assembly of the *R. hastatulus* XYY cytotype produced two sets of highly contiguous chromosome-level scaffolds. Haplotype A had an assembly size of approximately 1,510 Mb, with 95% of the genome assembled into four main scaffolds (Fig. 1; supplementary tables S1 and S2 and figs. S1 and S2, Supplementary Material online), which corresponds with the expected chromosome number for the X-bearing haplotype of three autosomes and one sex chromosome (Smith 1964; Rifkin et al. 2021). The BUSCO (Manni et al. 2021) completeness score was 99.3% (Eukaryota database) and 96.2% (Embryophyta database). Similarly, 97% of the haplotype B assembly was placed into the expected five main scaffolds (three autosomes and two Y chromosomes), and an assembly size of 1,719 Mb, 209 Mb larger than the haplotype A assembly (Fig. 1). The BUSCO completeness score was 99.6% (Eukaryota database) and 95.0% (Embryophyta database). The difference in assembly size between the two haplotypes is consistent with previous flow cytometry data, which indicated that the male genome is approximately 10% larger than the female genome (Grabowska-Joachimik et al. 2015). Cytological measurements suggest the two Y chromosomes combined are approximately 50% larger than the X/NeoX. These findings indicate substantial genome expansion has occurred on the Y chromosomes since they began diverging from the X (see below).

Our assembly of the hermaphroditic species *R. salicifolius* had a much more compact size of approximately 586 Mb, with 99.0% of the assembly found in the expected 10 scaffolds, based on chromosome counts of $x = 10$ (Löve 1986). The BUSCO completeness score was 99.6% (Eukaryota database) and 97.1% (Embryophyta database).

Using previously published transcriptome sequences from population samples of both males and females from the XYY cytotype (Hough et al. 2014), we were able to confirm the identification of the sex chromosomes in *R. hastatulus* and validate the high accuracy of the sex chromosome phasing (supplementary fig. S3, Supplementary Material online). In particular, we identified SNPs and insertion–deletion polymorphisms (indels) from a broad population sample that represent putative fixed differences between the X and Y chromosomes (all males heterozygous, all females homozygous for either the reference or alternative base). We found that 7,311 out of 7,333 fixed sex-specific SNPs and indels mapped to the largest scaffold (hereafter the X chromosome, approximately 483 Mb) of haplotype A, 7,281 (99.3%) of which had the female homozygous allele as the reference base. Similarly, 99.8% of fixed sex-specific SNPs and indels (6808/6823) mapped to two large scaffolds on haplotype B (hereafter Y1, 343 Mb and Y2, 348 Mb), and

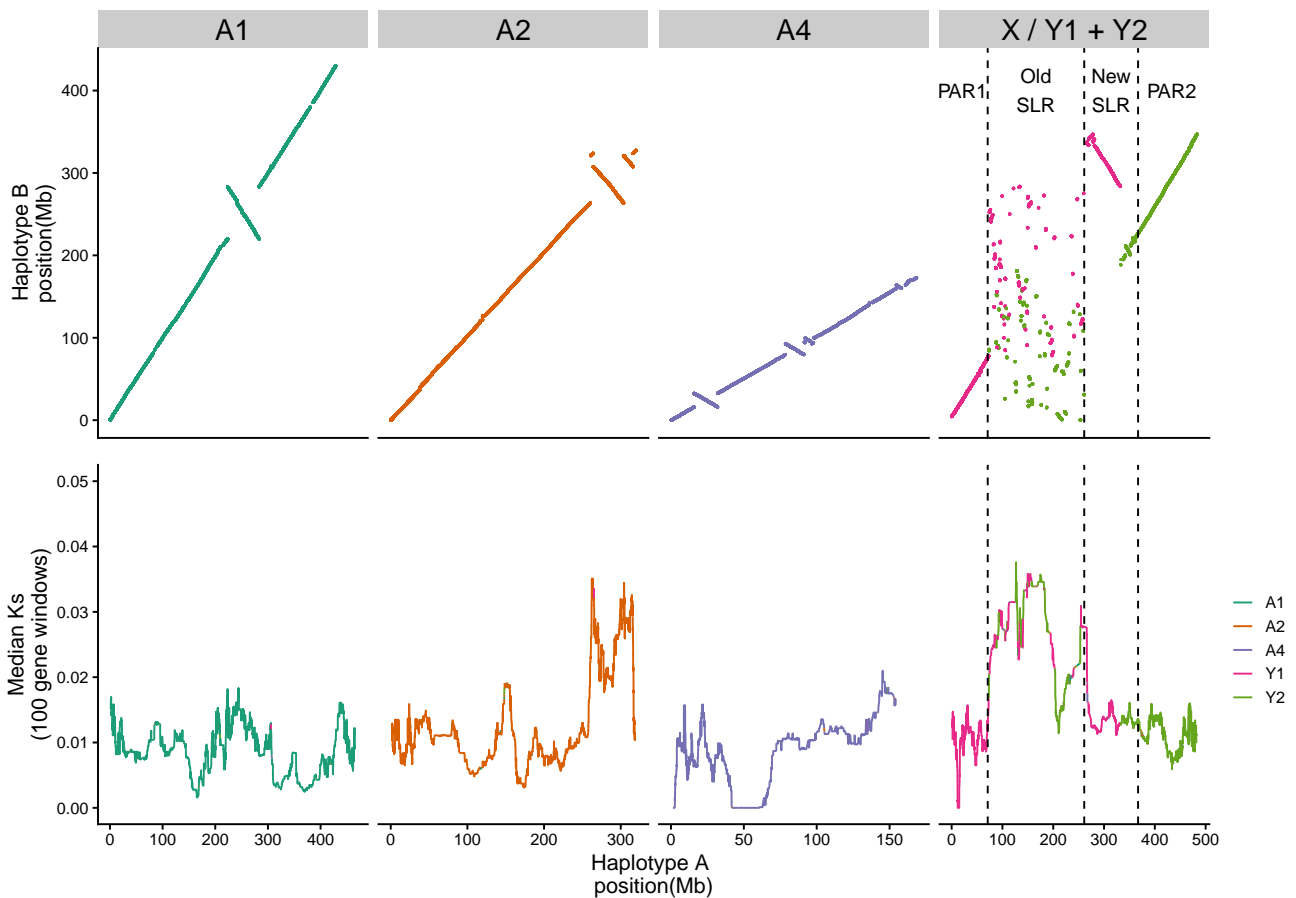


Fig. 1. Synteny and divergence between the two assembled haplotypes of XY male of *R. hastatulus*. Top panel: syntenic genomic position comparison based on whole-genome alignment for the autosomes and the sex chromosomes. Bottom panel: median Ks between syntenic genes in 100 gene windows with a step size of one gene along each chromosome, calculated from individual genes with Ks < 0.2. Autosomal terminology is used to remain consistent with genome assemblies from the XY cytotype. Autosome 3 (A3) from the XY cytotype is a component of the sex chromosomes in the XYY cytotype. The old sex-linked-region ('Old SLR') is shared with the XY cytotype; PAR1, a pseudoautosomal region, is shared with XY cytotype; the new sex-linked region ('New SLR') was formed from sex chromosome fusions; and PAR2, a recombining region, was formed from the neo-sex chromosome region.

99.9% of these fixed SNPs and indels contained the male-specific heterozygous allele as the reference.

To further validate the phasing of sex-linked regions more globally than at shared SNPs, we mapped short genomic reads from a male and female sample from this cytotype to a combined reference genome that included the autosomes and X chromosome from haplotype A along with the sex-linked regions of Y1 and Y2 from haplotype B. Male and female coverage across this assembly is as expected ([supplementary fig. S4, Supplementary Material online](#)); female genomic read coverage is greatly reduced on the Y chromosome, while male genomic read coverage is approximately halved on the sex-linked regions of the X and Y chromosomes compared with autosomes. These combined results highlight the high level of completeness and phasing accuracy of the assembled sex chromosomes.

Synteny Analysis

Whole-genome alignments integrated with syntenic gene anchors ([Song et al. 2022](#)) confirm a high level of synteny across the main autosomes (named according to the

naming conventions from the XY cytotype) between the two phased haplotypes of *R. hastatulus* ([Fig. 1](#)). However, several heterozygous large and small putative inversion differences are apparent across the three main autosomes, indicating a significant degree of inversion heterozygosity. Overall, eight putative heterozygous inversions could be identified on the autosomes, ranging in size from 189 kb to 39 Mb in length. These heterozygous inversions collectively span approximately 10% of the autosomes. Strikingly, three of these inversions, including two nested inversions on the second autosome (A2), show highly elevated levels of between-haplotype heterozygosity as measured by Ks in gene copies between the haplotypes ([Fig. 1](#)). Two of these inversions (the nested ones on A2) were independently identified in comparative genetic mapping between the two cytotypes ([Rifkin et al. 2021](#)), and these regions as well as the inverted region on A1 were identified as contributing divergent genotype clusters across populations within the XY cytotype ([Beaudry et al. 2022](#)). To further validate the largest putative inversions, we used the Omni-C contact map data ([supplementary fig. S5, Supplementary Material online](#)). Mapping of the Omni-C

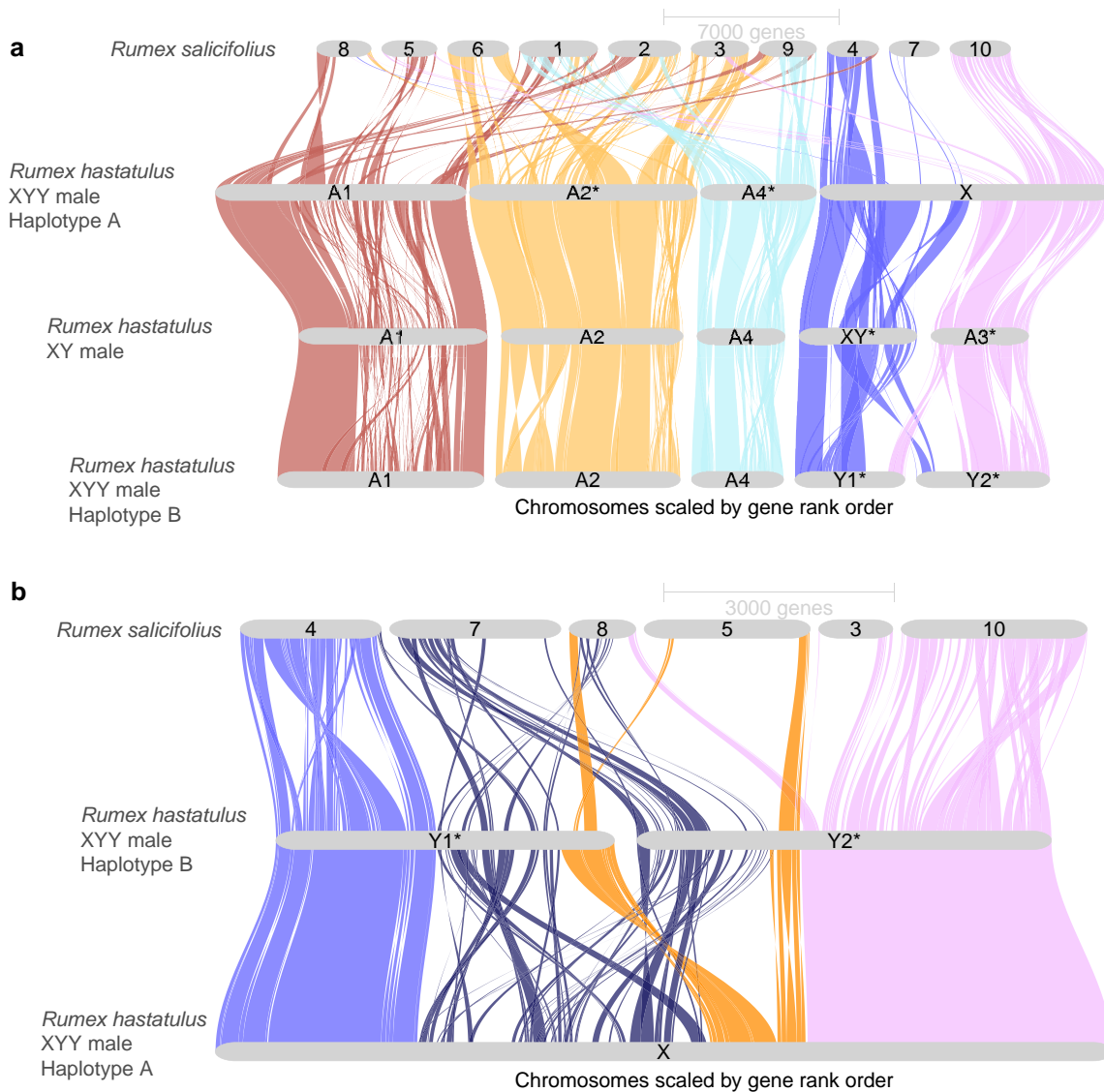


Fig. 2. GENESPACE riparian plots between assembled *Rumex* genomes. a) Syntenic gene blocks between the *R. hastatulus* XYY A and B haplotypes, *R. hastatulus* XY male, and *R. salicifolius*. b) Close-up view of synteny between the X and Y chromosomes of *R. hastatulus* XYY clade and orthologous regions in outgroup *R. salicifolius*. The pseudoautosomal regions are colored in light blue and pink (leftmost and rightmost regions on the X, respectively), old sex-linked region in dark blue (second region on the X), and new sex-linked region (third region on the X) in orange.

reads to the two separate haplotypes revealed evidence for long-range contacts for these largest inversions in both haplotypes. Since our heterozygous sample is being mapped to haploid assemblies in both cases and we see this long-range interaction to both haplotypes, this could be consistent with *bona fide* inversion heterozygotes that are being cross-mapped to the alternative haplotype. To investigate this further, we mapped the Omni-C reads to a combined reference genome that includes both haplotypes, and these long-range interactions are no longer apparent (supplementary fig. S5, Supplementary Material online), consistent with the expectations if these putative inversions are real. Taken together, these patterns suggest that a subset of these inversion polymorphisms have a deep coalescent time, are shared between the cytotypes, and may be subject to balancing selection, potentially due to spatially varying selection, as predicted by theory

(Kirkpatrick and Barton 2006) and as observed in several other taxa (Lowry and Willis 2010; Fuller et al. 2019; Todesco et al. 2020; Bieker et al. 2022).

In contrast with the autosomes, a large section of the sex chromosome shows almost no remaining large-scale synteny between the X and Y, highlighting that extensive chromosome rearrangements have occurred since the loss(es) of recombination (Fig. 1). Comparisons of the Y-bearing haplotype B assembly with the previously assembled XY cytotype genome (Fig. 2) and patterns of male-specific SNPs from the XY cytotype mapped to the new assembly (supplementary fig. S1, Supplementary Material online) reveal that both Y chromosomes contain segments of both the ancestral sex chromosome (“old sex-linked region”; Fig. 1) and much more syntenic segments of the neo-sex chromosomes recently derived from autosome 3 (“new sex-linked region”; Fig. 1), which recently formed the neo-X and neo-Y

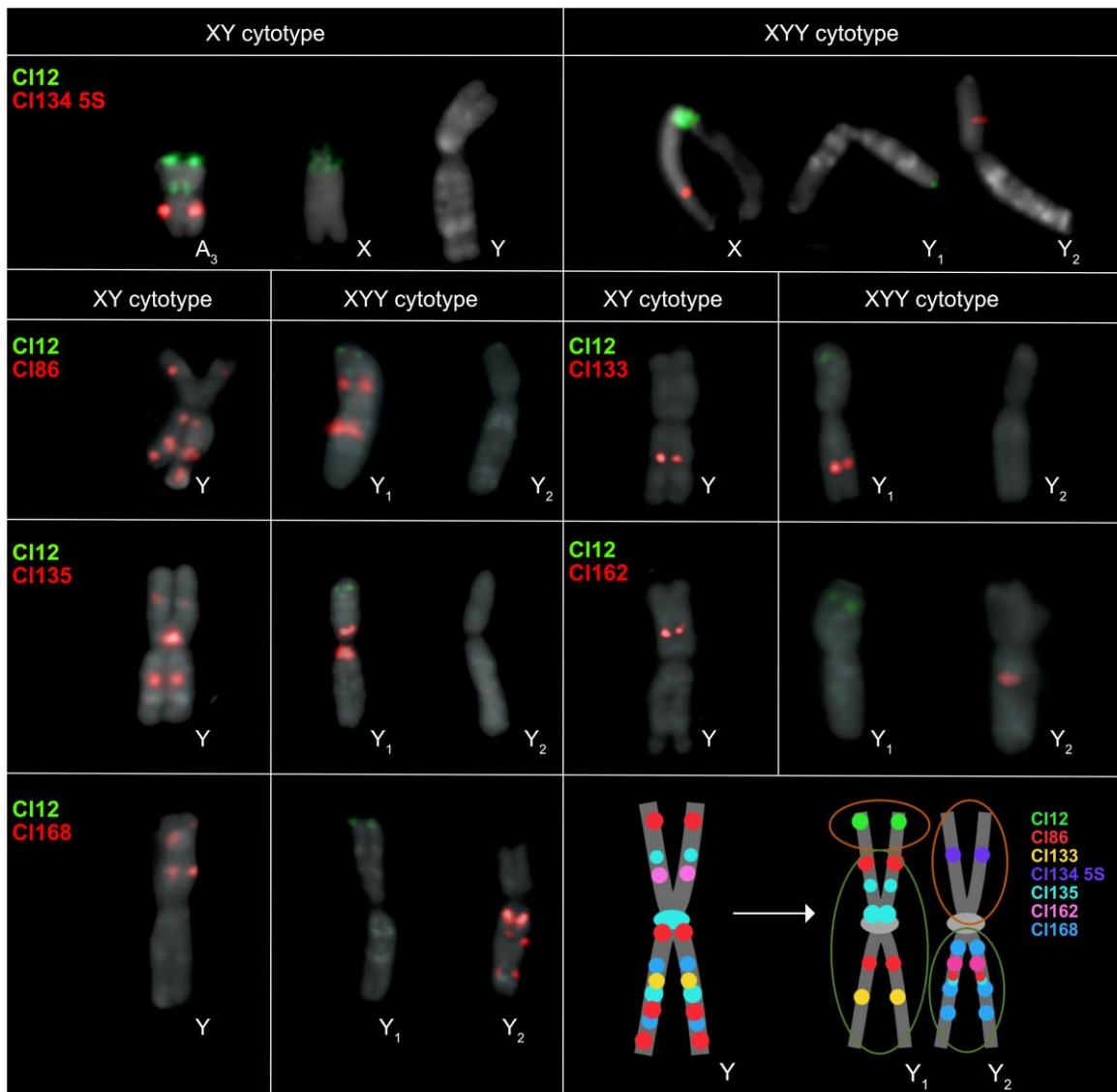


Fig. 3. Chromosomal location of seven tandem repeats on the metaphase chromosomes of XY and XYY cytotypes of *R. hastatulus*. The schematic depiction of the structure of the old Y, neo-Y₁, and neo-Y₂ chromosomes displays positions of all seven tandem repeats between Y chromosomes in both cytotypes. The position of CI12 and CI134 5S changes rapidly in the XYY cytotype, directly confirming translocation and fusion events between autosome 3 (in the XY cytotype) and the X-Y chromosomes. The orange (upper) circles represent the parts that originated from the autosome 3, and the green (lower) circles represent the old Y-chromosomal regions. Chromosome scale refers to [supplementary figs. S6 to S9, Supplementary Material](#) online.

chromosome regions. RepeatExplorer (Novák et al. 2013, 2020) analysis and cytogenetic mapping of seven sex-specific satellites (including CI134 5S together with CI12 that are located originally on autosome 3 in XY cytotype) in both cytotypes provided further support for the presence of both ancestrally autosomal regions and old sex-linked regions on both Y chromosomes, consistent with our scaffolding results (Fig. 3; [supplementary figs. S6 to S10, Supplementary Material](#) online). Further, the CI12 and its distribution on the neo-X chromosome suggest that the whole autosome 3 was fused together with the old-X ([supplementary fig. S6 to S10, Supplementary Material](#) online).

The patterns of fixed sex-linked SNPs from both cytotypes ([supplementary fig. S3, Supplementary Material](#) online) confirm the presence of a massive sex-linked region

(Fig. 1), spanning approximately 297 Mb on the X chromosome and 503 Mb on the Y chromosomes. The absence of sex-limited SNPs at the tips combined with previous comparative genetic mapping results (Rifkin et al. 2021) and early cytogenetic work (Smith 1964) suggests that the sex chromosomes have two pseudoautosomal regions, one on either side of the large, fused X (Figs. 1 and 2), where Y₁ retains the pseudoautosomal region from the ancestral Y (PAR1) and Y₂ contains a pseudoautosomal region derived from the ancestral autosome (PAR2). Altogether these results indicate that, in addition to the X-autosome fusion event, a secondary reciprocal translocation occurred between the homologous autosome and the ancestral Y chromosome. This additional translocation was previously hypothesized from cytological

data (Smith 1964) and may have been important to stabilize meiotic pairing, as shown for *Rumex acetosa* Y1XY2 trivalent structure during pachytene synapsis (Cunado et al. 2007). The difference in outcomes of the reciprocal translocations on the X and Y likely stems from an inversion on the ancestral autosome before or after the fusion with the X or the translocation with the Ys, as there is no evidence of loss of gene segments on either the neo-X or the neo-Y segments (Fig. 2). This is further supported by fluorescence in situ hybridization (FISH) results, which show that all main repeat clusters from the ancestral autosome are found on the neo-X, with evidence of several paracentric and at least one pericentric inversion event on both neo-Ys, further supporting our synteny analysis (Fig. 1). Multiple inversion events even between the cytotypes in the old Y-linked regions are evident from the new localization of satellite clusters Cl86, Cl133, Cl135, Cl162, and Cl168 (supplementary figs. S3 and S10, Supplementary Material online). It is possible for such large chromosomal rearrangements to occur in a single catastrophic event, as hypothesized for single chromosome shattering in the *Camelina* genome (Mandáková et al. 2019). On the other hand, the satellite enrichment in the XY cytotype could allow for such reorganization, given the new satellite and genome order on the neo-Ys.

In the neo-sex-linked regions, synteny is much more retained on this young sex chromosome pair (Fig. 1). However, four inversions are apparent within this stretch of approximately 102 Mb of new sex-linked sequence, capturing 31% of the region in heterozygous inversions, considerably higher than observed on the autosomes (eight inversions capturing 10% of the sequence in approximately 1 Gb of the genome). Note that inspection of contact maps in the combined reference mapping showed no evidence of spurious assemblies in these regions. These findings suggest that the recent formation of the neo-sex chromosomes and loss of recombination is accompanied by an elevated maintenance and/or high rate of spread of inversions following the chromosomal fusions.

Comparisons of syntenic gene order in hermaphroditic *R. salicifolius* indicate that, while there have been massive rearrangements genome wide (Fig. 2a), synteny breakdown has been much more extensive on the Y chromosome compared with the X in the sex-linked region (Fig. 2b). Specifically, we identify 155 orthologous genes where *R. salicifolius* and the old X chromosome retain syntenic positions whereas the Y position is nonsyntenic, and only 13 cases where the old Y and *R. salicifolius* have retained their positions to the exclusion of the X (supplementary table S3, Supplementary Material online). This excess is much greater than the relative difference in nonsyntenic orthologs on the autosomes of the two haplotypes (contingency test $\chi^2 = 26.183$, $df = 1$, $P < 0.001$). Interestingly, the pseudoautosomal regions appear to be derived mostly from different ancestral chromosomal origins than the sex-linked regions (Fig. 2b). This inference is in line with other chromosomes, where central regions of the chromosome that are associated with large regions of

very low recombination (Rifkin et al. 2022) appear to often have been derived ancestrally from different chromosomal regions than the arms, assuming *R. salicifolius* is closer to the ancestral state. The old sex-linked region derives primarily from two *R. salicifolius* chromosomes, scaffolds 7 and 8. To explore whether these two distinct segments represent evolutionary strata that were added to the sex-linked region at distinct times since the formation of the sex-linked region, we estimated K_s between X and Y-linked gametologs, the per nucleotide synonymous substitution rate for each sex-linked gene. We found no evidence for a significant difference in the number of “young” ($K_s < 0.03$) relative to “old” ($K_s > 0.03$) sex-linked genes derived from the two *R. salicifolius* chromosomes (chi-square contingency test, $\chi^2 = 3.0634$, $df = 1$, $P = 0.08$). Furthermore, while there is heterogeneity across the X chromosome in median X-Y divergence, there is no clear evidence of discrete “evolutionary strata” involving distinct chromosomal segments in the old sex-linked region (Fig. 1). This finding may be due to the extensive chromosomal rearrangements that have occurred since the origins of the sex-linked region, the origins of the sex-linked region from a preexisting region of reduced recombination without strata and/or an ongoing history of gene conversion between some sex-linked genes.

Genomic Distribution of Repeats

Previous work indicated that all *R. hastatulus* chromosomes have large, repeat-rich regions of low recombination, including the sex-linked regions (Rifkin et al. 2021, 2022). A resulting question is whether further loss of recombination on the sex-linked regions of the Y chromosomes drives additional and distinct repeat accumulation. As expected given the genome size differences, *R. hastatulus* (84% and 86% on haplotypes A and B, respectively) has more TEs overall than *R. salicifolius* (66.41%) (Fig. 4; supplementary fig. S11, Supplementary Material online). Despite the high levels of repetitive content genome wide in *R. hastatulus*, repeat annotation of our phased assemblies reveals that the Y chromosomes have considerably more TEs than the X or autosomes (Fig. 4; supplementary fig. S12, Supplementary Material online). Mutator-like DNA elements show a major localized accumulation on Y1 and a minor accumulation on Y2, copia-like elements show additional accumulation on Y2, and Ty3 elements have accumulated in localized positions on both Y1 and Y2 (Fig. 4; supplementary fig. S13, Supplementary Material online). The older sex-linked regions of both Y1 and Y2 have higher repeat content than the older sex-linked regions of the X (Fig. 4b). Overall copy number is significantly elevated by almost 3-fold on the old sex-linked Y region compared with the old sex-linked X (supplementary table S4 and fig. S14b, Supplementary Material online; chi-squared test, $P \ll 0.001$). In contrast, TE copy number is marginally elevated (1.09-fold) on the newly sex-linked region of the X compared with the Y, but this is only slightly higher than the difference between the PARs (1.02-fold) (supplementary table S4, Supplementary

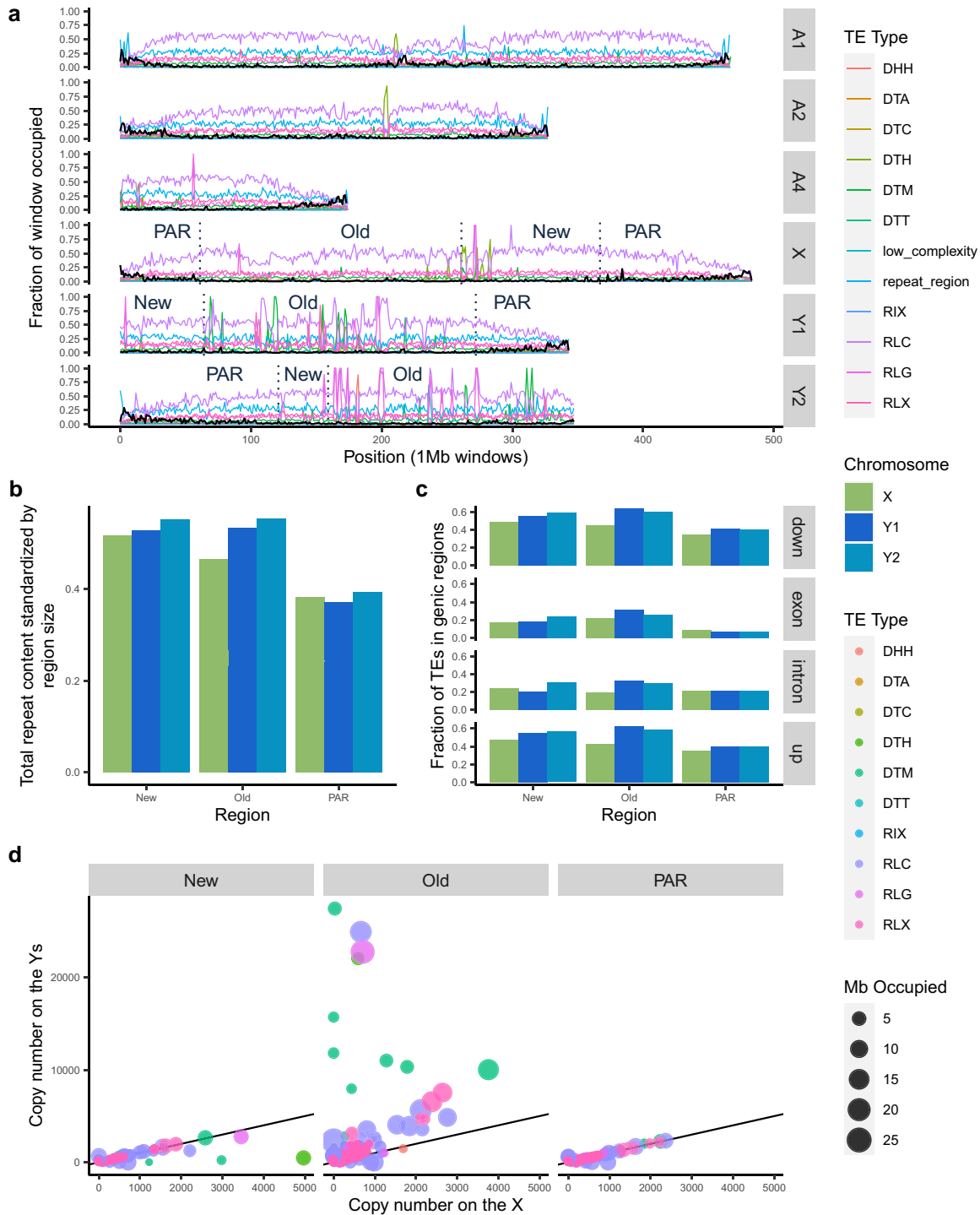


Fig. 4. TE distribution across the genome of *R. hastatulus* indicates increased TE content on the Y chromosomes relative to the X chromosome or autosomes. a) Proportion of 1-Mb nonoverlapping windows occupied by genes and TE sequences. Values for autosomes and Y chromosomes are from haplotype B, except for the values shown on the X, which are from haplotype A. Black lines are the fraction of the window that is genic sequence and colored lines are fractions of the window that is TE sequence. The dotted vertical lines and labels on the sex chromosomes divide the chromosomes into newly sex-linked (New), old sex-linked (Old), and pseudo-autosomal regions (PAR). b) The summed length of annotated repeats on the newly sex-linked, old sex-linked, and pseudoautosomal regions of the sex chromosomes divided by the size (total bp) of the region. Considerable nesting and overlap of repeats have led to ratios larger than one. c) For each gene on the sex chromosomes, the average proportion of nearby TEs within the newly sex-linked, old sex-linked, and pseudoautosomal regions. The genic regions were separated into, in order from top to bottom: 1-kb upstream, within exons, within introns, and 1-kb downstream. The X genes are LiftOff matches from haplotype B and the Y genes were filtered for those that LiftOff found a match for in haplotype A. d) Each point represents a family of TEs. The size of the point reflects the number of Mb occupied by members of the family (includes all sex chromosomes) and the X and Y axis values reflect the number of family members on the X and Y chromosomes, respectively. The black line has a slope of 1, to visually indicate where equal quantities would fall. The three-letter transposon codes in the legend are from [Wicker et al. \(2007\)](#). DHH: Helitron; DTA: hAT; DTC: CACTA; DTH: Harbinger; DTM: Mutator; DTT: Tc1-Mariner; RIX: unknown LINE (long interspersed nuclear element); RLC: Copia; RLG: Ty3; RLX: unknown LTR.

Material online). Given its similarity to the difference in PARs, the difference between the newly sex-linked regions may reflect stochastic differences between haplotypes and minor technical differences in TE annotations.

Transposon families are a useful unit of comparison for understanding TE abundance in the two haplotypes. Wicker et al. (2007) proposed an 80–80–80 rule of similarity to group transposons into families. The procedure requires that the TEs be at least 80 bp in length and have 80% similarity over 80% of the aligned sequences. PanEDTA uses this definition to group the annotated TEs into families across the two haplotypes, which allows for a more direct comparison of TE complement. Many individual TE families occupy more space and are more numerous on the older sex-linked regions of the Y chromosomes relative to the X (Fig. 4d). This pattern is especially true for harbinger, mutator-like, and long terminal repeat (LTR) elements (supplementary fig. S14, Supplementary Material online). Some of this accumulation has led to extreme clusters of very high copy numbers on the old Y, suggestive of local targeted transposition and/or expansion via tandem arrays (supplementary fig. S8, Supplementary Material online).

These results suggest extensive accumulation of TEs has occurred on the older sex-linked regions of the Y chromosome, but it is unclear whether this accumulation may be affecting genes. To understand whether this TE accumulation is primarily occurring in already repeat-dense areas, the overlap between the TE annotation and gene annotation was examined. To make comparisons as equivalent as possible for this analysis (given potential differences in the outcome of gene annotation due to differences in repeat content and other factors), we used a gene liftover (see Materials and Methods) of the haplotype B genome annotation to the haplotype A genome annotation and only retained genes with at least one open reading frame in both the original and lifted over annotation. Since the gene models were retained from the Y-bearing haplotype, this should be conservative with respect to our test for additional insertions near genes, since our filtered gene models from this haplotype should be biased against having TE insertions.

We observed significantly elevated numbers of TEs inside and near genes on the Y chromosomes, particularly in the old sex-linked region (Fig. 4c; supplementary table S4, Supplementary Material online). In contrast, genes in neo-sex-linked regions showed no signs of rapid TE accumulation on the Y, as differences between X and Y are similar to baseline differences between the PARs (Fig. 4c; supplementary table S4, Supplementary Material online). Overall, we found signs of considerable accumulation of TEs in our older sex-linked region of the Y chromosomes, including into and near genes, although to a lesser extent than TE accumulation further from genes.

Gene Retention and Loss

Previous studies of gene loss using transcriptome and short-read genome information on plant sex chromosomes have focused on the pairwise comparison of X and Y

chromosomes (Hough et al. 2014; Bergero et al. 2015; Papadopulos et al. 2015; Beaudry et al. 2017; Crowson et al. 2017). This approach cannot distinguish between gene loss and gene movement or duplication among sex chromosomes and autosomes. The genome of a hermaphroditic outgroup, in this case our *R. salicifolius* assembly, allows for the specific identification of genes not present on one of the *R. hastatulus* sex chromosomes that were “ancestrally” present in the same syntenic block. This in turn enables quantification of the extent of *bona fide* gene loss on the sex chromosomes by identifying syntenic orthologs in the outgroup.

Compared with all autosomes and the neo-sex chromosome, there is a high proportion (~34%) of genes in the old sex-linked region that show evidence of loss on the Y chromosome despite their syntenic presence in both *R. salicifolius* and the X chromosome (Fig. 5a; supplementary table S5, Supplementary Material online). Approximately 38% of the lost genes still showed fragments on the Y chromosome and were classified as partially lost (defined by less than 50% of the putatively missing gene with similarity to the Y chromosome), whereas the remainder are inferred to be fully deleted. These estimates are much higher than on autosomes or the X chromosome, suggesting that the extent of loss is much greater than expected simply from gene copy number variation and/or bioinformatic errors. Overall, if we use the autosomal “loss” values as a baseline for the presence–absence polymorphism and/or technical error, we see approximately 30% of genes have been lost on the Y chromosome in the old sex-linked region. Patterns of gene loss along the Y chromosome show evidence of regional variation in the extent of loss, particularly when anchored to the *R. salicifolius* genome with a likely more ancestral gene order (Fig. 5b and c). This finding could reflect either the presence of large-scale regional deletions and/or a dynamic history of recombination suppression (i.e. evolutionary strata).

In contrast, we see no sign of excess gene loss on the old X-linked region (supplementary table S6, Supplementary Material online; Fig. 5a), providing no evidence of early gene loss on the X chromosome, as found recently in other systems (Mrnjavac et al. 2023). Furthermore, there is no sign of excess gene loss in the “new” sex-linked region (NeoY), suggesting a lack of rapid deletion of Y-linked genes since the chromosomal fusion. Among the genes lost in the neo-X and autosomes, almost all are classified as partially lost. In particular, the evidence for complete gene loss of syntenic orthologs is nearly exclusively restricted to the old Y (159 genes fully lost on the Y, compared with only 24 completely lost in the rest of the genome).

Conclusions

Our results provide two time points early in the evolution of heteromorphic sex chromosomes, supported by a hermaphroditic outgroup. Our studies revealed that in the extremely young (<200,000 generations) neo-sex linked regions of

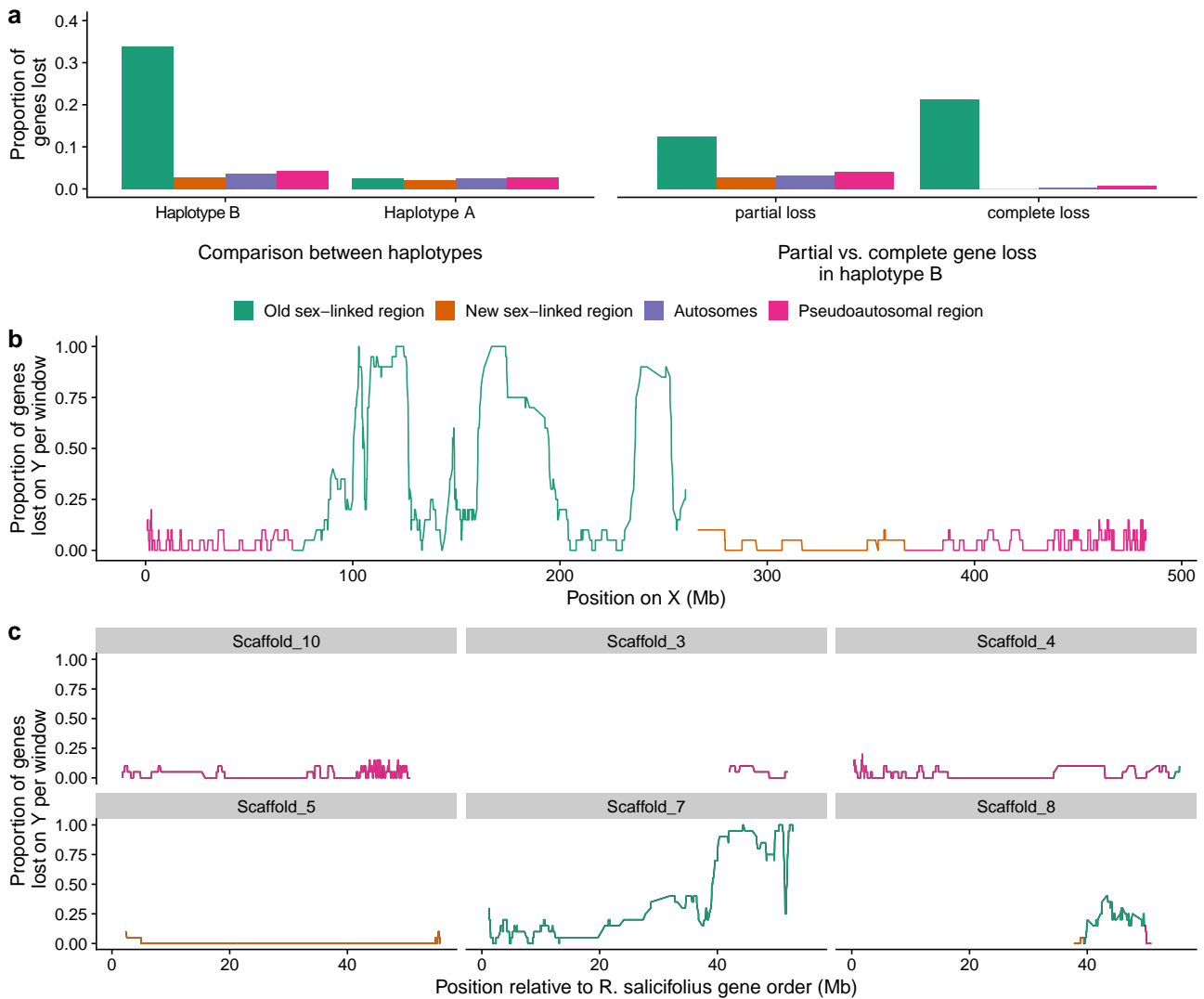


Fig. 5. a) Proportion of genes lost within the sex-linked regions, pseudoautosomal regions, and autosomes in *R. hastatulus*. Left panel: comparison of genes that are absent from the Y-bearing haplotype B, while present on the X-bearing haplotype A assembly, and vice versa. Right panel: proportion of partially lost versus completely lost genes across regions. b) Proportion of genes lost on Y calculated in 20 gene windows along the X chromosome. c) Proportion of genes lost on Y calculated in 20 gene windows with respect to the positions of their syntenic genes in *R. salicifolius* scaffolds.

R. hastatulus, chromosome rearrangements have accumulated rapidly without signs of gene loss or TE invasion. In contrast, on the older (but relatively young, <10 MYA) regions of the sex chromosomes, extensive rearrangements have led to a near-complete breakdown of synteny, TE invasion, and extensive gene loss. The extent of rearrangement is striking for a relatively young sex chromosome system that retains low X-Y divergence for many of the genes that remain. This extent of Y degeneration and sex chromosome evolution is in line with recent results from an unrelated dioecious plant, *Silene latifolia*, which has an approximately 11 MYA Y chromosome and includes strata as young as 5 MYA (Moraga et al. 2023; Akagi et al. 2023). The emergence of sex-linked regions in large pericentromeric regions of low recombination may contribute to a highly dynamic genetic system that has evolved heteromorphic sex chromosomes over a relatively short time period.

Materials and Methods

Long-Read Genome Sequencing

We grew a male and female plant from two independent maternal families of *R. hastatulus* from the XYY clade collected from Marion, South Carolina (Pickup and Barrett 2013) in the University of Toronto glasshouse. Following full-sib mating from this F₁ generation, 11 g of leaf tissue from a single F₂ male was sampled to extract high-molecular-weight DNA conducted by Dovetail Genomics (Cantata Bio, LLC, Scotts Valley, CA, USA). A total of 4,618,456 PAC Bio CCS reads (Pacific Biosciences Menlo Park, CA, USA) were sequenced by Dovetail for a total of 87.7 Gb (approximately 46× coverage, based on a male genome size estimate of 1.89 Gb, Grabowska-Joachimik et al. 2015). Similarly, we ordered a single *R. salicifolius* plant from seed collected from Nevada, USA from the United States Department of

Agriculture's US National Germplasm System (Accession RUSA-SOS-NV030-372-10) and collected 20 g of leaf tissue for high-molecular-weight DNA extraction and sequencing. A total of 5,149,926 PAC Bio CCS reads were sequenced totaling 75.3 Gb (approximately 108× coverage based on our flow cytometry estimate of 696 Mb).

PacBio Library and Sequencing

DNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20 kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) using the manufacturer's recommended protocol. The library was bound to polymerase using the Sequel II Binding Kit 2.0 (PacBio) and loaded onto PacBio Sequel II. Sequencing was performed on PacBio Sequel II 8M SMRT cells.

Dovetail Omni-C Library Preparation and Sequencing

Proximity ligation and sequencing was conducted by Dovetail using Omni-C sequencing for both species. For each Dovetail Omni-C library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNase I, and chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed, and the DNA was purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeqX platform to produce an approximately 30× sequence coverage.

De Novo Assembly

For the male *R. hastatulus* sample, we conducted a haplotype-resolved de novo assembly using Hifiasm v. 0.16.1-r375 (Cheng et al. 2021), using the Omni-C sequencing for haplotype resolution. Paired-end Omni-C reads were then mapped and filtered to the two phased assemblies using bwa v0.7.15 (Li and Durbin 2009) following the Arima mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline), and resulting filtered (MapQ > 10) bam files had duplicates marked using Picard v2.7.1. We scaffolded both haplotypes of the assembly using YAHS v1.2a.2 (Zhou et al. 2023) to generate scaffolded assemblies from each phased haplotype. We manually inspected the scaffolded assembly using a combination of Juicebox v1.11.08 (Durand, Robinson, et al. 2016) and whole-genome alignment to our previous assembly from the XY cytotype (Rifkin et al. 2022) to identify and break one false join in the assembly. In particular, a break was inserted at the point between autosome 4 and Y2 in haplotype B based on manual inspection. The X-bearing haplotype assembly is referred to as haplotype A, while

the Y-bearing haplotype assembly is referred to as haplotype B. Each haplotype contains one copy of each autosome of which parental origins are unknown. Quality control was performed using NCBI's foreign contamination screening tool FCS-GX (Astashyn et al. 2024), and sequences flagged as contaminants were removed.

For *R. salicifolius*, Hifiasm v0.15.4-r347 was run by Dovetail to generate the primary contigs. Because this is a hermaphroditic species, we opted for the primary assembly option to generate a comprehensive/best quality contig-level assembly, without a full phasing of the genome. BLAST (Altschul et al. 1990) results of the *R. salicifolius* Hifiasm output assembly against the nt database were used as input for BlobTools v1.1.1 (Laetsch and Blaxter 2017), and scaffolds identified as possible contamination were removed from the assembly. Finally, purge_dups (Guan et al. 2020) v1.2.5 was used to remove haplotigs and contig overlaps.

The primary assembly was scaffolded by Dovetail using the Omni-C reads with the HiRise assembler (Putnam et al. 2016), after aligning the Omni-C library reads to the filtered draft input assembly using bwa v0.7.15 (Manni et al. 2021).

The separations of Dovetail Omni-C read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, score prospective joins, and make joins above a threshold.

Assembly completeness was assessed using BUSCO v5.4.4 (Manni et al. 2021) using both the *Embryophyta* and *Eukaryota* databases.

Contact Maps

To construct contact maps for *R. hastatulus* and *R. salicifolius*, we used bwa v0.7.15 to align PacBio reads to our final assembly. We then used pairtools v1.0.2 to create a pairsam file with annotations of ligation events and potential pairs (Open2C et al. 2023). The minimum threshold used for defining multimapping alignments was 40 and the maximum gap between alignments was 30. We then sorted the parsed pairsam file and marked duplicate pairs using pairtools (pairtools sorted and dedup). The pairs were then split into a bam and pairs file using pairtools split. Lastly, we converted the pairs to contact maps using Juicer's precommand (Durand, Shamim, et al. 2016). All contact maps were visualized using the Juicebox visualization environment (Durand, Robinson, et al. 2016).

Sex-Linked SNP Identification

We mapped RNAseq leaf expression data from population samples of TX and NC cytotype male and female plants of *R. hastatulus* (Hough et al. 2014) to both haplotype assemblies using STAR v2.7.10a (Dobin et al. 2013). We performed variant calling using freebayes v1.3.4 (Garrison and Marth 2012) and then filtered sites to a final set comprised of biallelic sites with genotype quality > 30. We then used custom R scripts to identify putative sex-linked SNPs.

We selected all sites that were heterozygous in all six males and homozygous in all six females per cytotype to obtain candidate fixed SNP differences between X and Y.

Coverage of Genomic Reads

Genomic reads from an XYY male and female F2 individuals resulting from two full-sib crosses of seed originally from Marion, South Carolina (Pickup and Barrett 2013). These individuals were not from the same crosses as the sample used for long-read sequencing. These individuals were sequenced to 20× coverage using the Illumina NovaSeq platform and mapped using bwa v0.7.15 to a merged version of the diploid assembly that contains one copy of each autosome and PARs, as well as the X and Y chromosomes. Genomic read coverage was evaluated using the Qualimap v.2.3 bamqc function, which calculates mean coverage in 400 windows across the genome (Okonechnikov et al. 2016).

Gene Annotation

Gene annotation followed previous approaches (Rifkin et al. 2022). For *R. hastatulus*, we performed the annotation with MAKER-3.01.03 (Cantarel et al. 2008) in four rounds. In the first round, we used the *R. hastatulus* RNA-seq transcripts from previously published floral and leaf transcriptomes (Hough et al. 2014; Sandler et al. 2018) and annotated Tartary buckwheat proteins from version FtChromosomeV2.IGDBv2 (Zhang et al. 2017) for inferring gene predictions and used the TE library (see below) to mask the genome. We trained the resulting annotation for SNAP gene predictor, using the gene models with an AED of 0.5 or better and a length of 50 or more amino acids. In the following rounds, we used the resulting EST and protein alignments from the first round and the SNAP model from the previous round for annotation. We functionally annotated the final gene models based on BLAST v2.2.28+ (Altschul et al. 1990) and InterProScan 5.52 to 86.0 (Jones et al. 2014), by using the related scripts in the Maker package. For *R. salicifolius*, we used the same approach, except we integrated RNAseq data for this species from flower buds, pollen and leaves (Hibbins et al. 2023) and a TE library generated from Repeat Modeller (Smit and Hubley 2008).

Syntenic Gene Alignments and Analysis

We estimated orthology and synteny between protein-coding genes in haplotype A, haplotype B, and *R. salicifolius* using the R package GENESPACE v1.1.8 (Lovell et al. 2022), which uses MCSanX (Wang et al. 2012) to infer syntenic gene blocks and then implements ORTHOFINDER v2.5.4 (Emms and Kelly 2019) and DIAMOND v2.1.4.158 (Buchfink et al. 2021) to find orthogroups within syntenic blocks. We performed analyses and visualized results in Rv4.1.0 (R Core Team 2022). We used default parameters, with the exception of ORTHOFINDER one-way sequence search, which is appropriate for our closely related genomes.

We also conducted whole-genome pairwise alignments between the two haplotypes using AnchorWave v1.01

(Song et al. 2022), using the options allowing for relocation variation, and chromosome fusion. We used Minimap2 (Li 2018) in the AnchorWave alignment, followed by Proalign using “-Q 1” option.

Ks Analysis

We calculated synonymous substitution rate between haplotype assemblies A and B using SynMap2 on the COGE platform (Haug-Baltzell et al. 2017). To compare homologous genes between haplotypes, we used a cutoff of $K < 0.2$. We plotted median Ks values in 100 gene sliding windows (step size = 1) relative to their positions on the X chromosome

Gene Gain and Loss

Pangenome annotations produced by GENESPACE provide a list of orthologous genes shared by each genome and their positions relative to an assigned reference genome. We excluded genes with nonsyntenic orthologs and genes belonging to arrays that were not defined as representative by GENESPACE from subsequent analysis. We calculated the number of genes lost on haplotypes A and B by counting the number of syntenic genes found in both *R. salicifolius* and the other phased haplotype but absent from the focal haplotype assembly. To determine whether candidate lost genes are indeed lost and not simply missing from the annotation, we performed BLAST v2.5.0+ (Altschul et al. 1990) of these gene transcripts to the entire genome assembly sequence. We selected only the top BLAST hit (by percent identity) per candidate lost gene. We classified genes as present if the top BLAST sequence was on the corresponding chromosome. We classified genes not meeting this condition as lost, as well as genes where less than 50% of the query sequence is aligned to the subject. We defined these genes with less than 50% of query aligned as partially lost and included them within the total number of lost genes.

Nonsyntenic Orthologs

We identified one-to-one orthologs within the pangenome annotation where a syntenic ortholog was shared with the outgroup, *R. salicifolius*, in only one of the haplotypes, while the other haplotype's ortholog was nonsyntenic, as defined by GENESPACE. To determine whether there is an association between sex-linked regions and haplotype in terms of nonsyntenic ortholog content, we performed a 2×2 chi-square test of independence (R v4.3.1) comparing counts in the old sex-linked region and all autosomes for both haplotypes.

Satellite Identification, TE Annotation, and Analysis

We identified satellites using RepeatExplorer2 (Novák et al. 2020). We then preprocessed short-read Illumina data (Beaudry et al. 2017) with RepeatExplorer's inbuilt preprocessing pipeline (Novák et al. 2013, 2020). Trimming step was set to keep only full-length 150-bp reads and discard low-quality reads (quality cutoff = 10, percent above cutoff = 95) or

reads containing adapters. We ran RepeatExplorer2/TAREAN pipeline (v0.3.8 to 451; [Novák et al. 2017, 2020](#)) with default parameters. To compare repeats in the four samples, a comparative analysis was run following [Novák et al. \(2020\)](#), where we analyzed reads from all samples together with equal coverage between samples (using genome sizes according to [Grabowska-Joachimik et al. 2015](#)). We further used sex and cytotype-specific clusters for cytogenetic analysis ([supplementary table S7, Supplementary Material online](#)).

We produced the TE annotation using EDTA (Extensive de-novo TE Annotator) v2.1.0 pipeline ([Ou et al. 2019](#)). This pipeline combines the best-performing structure- and homology-based TE finding programs (GenomeTools, LTR_FINDER_parallel [[Ou and Jiang 2019](#)], LTR_harvest_parallel [[Ellinghaus et al. 2008](#)], LTR_retriever [[Ou and Jiang 2018](#)], Generic Repeat Finder [[Shi and Liang 2019](#)], TIR-Learner [[Su et al. 2019](#)], HelitronScanner [[Xiong et al. 2014](#)], and TESorter [[Zhang et al. 2022](#)]) and filters their results to produce a comprehensive and nonredundant TE library. We used the optional parameters “-sensitive 1” and “-anno 1” to identify remaining unidentified TEs with RepeatModeler and to produce an annotation. We used custom R scripts to visualize the data.

To analyze insertions near genes, we used Bedtools v2.30.0 and custom R scripts to compare the TE annotation file against the gene annotation, using genes lifted over from haplotype B to haplotype A with LiftOff 1.6.3 ([Shumate and Salzberg 2021](#)).

Chromosome Preparation and Cytogenetic Analysis

We used young seedlings of *R. hastatulus* of both XY, XYY cytotypes (North Carolina and Texas) for chromosome preparation, cell synchronization, and metaphase chromosome arrest as described in [Bačovský et al. \(2020\)](#). Additionally, we grew plants of XYY cytotype in a hydroponic tank with Hoagland solution in a growth chamber with a 16-h light/8-h dark cycle at 22 °C ([Hoagland and Snyder 1933](#)). We collected young roots from hydroponic tanks once per every 2 wk, synchronized in ice-cold water at 4 °C for 24 to 28 h. After the cell synchronization, we immediately fixed the root tissue in Clarke's fixative (ethanol:glacial acetic acid, 3:1, v:v) and stored it at 37 °C. After 1 wk of fixation, we replaced the fixative, and fixed roots were stored at -20 °C until further use.

We isolated the DNA used for PCR amplification from young leaves using cetyltrimethylammonium bromide (CTAB) solution and chloroform. We ground young leaves in a sterile grinder in liquid nitrogen and added 1 ml of CTAB solution to each sample. We vortexed the mixture for 30 s and incubated it at 65 °C for 45 min. We then added 2 µl of RnaseA (concentration 200 mg/ml) for the last 5 min. Next, we added 700 µl of chloroform and acetic acid solution (chloroform:glacial acetic acid, 24:1, v:v) to each sample, vortexed each for 1 min and centrifuged at maximum speed (14,000 rpm) for 2 min. We transferred the upper aqueous layer to a new tube and added 700 µl of chloroform and then vortexed and centrifuged the

samples for an additional 5 min at 14,000 rpm. We again transferred the upper aqueous layer to a new tube and precipitated the resulting DNA with 800 µl of isopropanol. We vortexed and centrifuged the mixture at maximum speed for an additional 5 min, then discarded the supernatant, and repeated the whole step using 75% ethanol to remove any excess salts. Finally, we air-dried the pellet for 5 min and dissolved in 20 to 40 µl of 1× Tris-EDTA buffer for 45 min. We then analyzed isolated DNA on 1% agarose gel, and its concentration and purity were measured on Nanodrop.

Based on the RepeatExplorer analysis ([supplementary fig. S6, Supplementary Material online](#)), we designed new primers for XY/XYY cytotype and sex-specific satellites in GeneiousPrime (2023.1.1) and synthesized in GeneriBiotech. We used primers directly for PCR amplification ([supplementary table S7, Supplementary Material online](#)), and we amplified the satellites to the manufacturer's instructions using 0.4 µl of *R. hastatulus* gDNA (TopBio, Vestec, Czech Republic, T034). The PCR conditions were as follows: 4 min at 94 °C, 36 cycles of 20 s at 94 °C, 20 s at 50 to 60 °C ([supplementary table S7, Supplementary Material online](#)), 30 s at 72 °C (for CI135 1 min), and final extension 5 min at 72 °C. We verified the PCR products by 1% agarose electrophoresis with EtBr staining and purified using the QiaQuick PCR Purification Kit (Qiagen, Hilden, Germany, 28104) following the manufacturer's instructions. We then verified the purified products again by agarose electrophoresis, followed by Nanodrop measurement.

We labeled the purified DNA by nick translation according to the manufacturer's instructions using Atto488 NT (PP-305L-488), Atto550 NT (PP-305L-550), or Cy5 (PP-305L-647N) (Jena Bioscience, Jena, Germany). The reaction proceeded for 1 h and 30 min at 15 °C. We verified the nicked-DNA products on 1% agarose gel with EtBr staining. We then placed the reactions on ice directly after the reaction to avoid the overlabeling of DNA before addition of EDTA. We used the nick-translated products as DNA probes in FISH. The hybridization mixture (87% stringency; [supplementary table S8, Supplementary Material online](#)) included 1 µl of labeled DNA (the final volume 1.5 ng/µl). We carefully mixed the hybridization mixture and denatured the sample at 85 °C for 10 min and transferred on ice for 5 min to perform FISH.

We prepared chromosomes using the squashing technique as described in [Karafiátová et al. \(2016\)](#) and [Bačovský et al. \(2020\)](#) with minor modifications, using 0.05 M HCl acid in 0.001 M citrate buffer before enzymatic digestion. We used slides containing chromosomes with well-preserved morphology and structure for FISH, as described in [Schubert et al. \(2016\)](#) with minor modifications. Briefly, we first washed the slides for 2× 5 min in 2× SSC solution (pH 7.2 to 7.5), then refixed the slides in Clarke's fixative for 10 min, and washed 2× 5 min in 2× SSC solution. To remove the remnants of cytoplasm, we treated slides with pepsin (50 mg/ml) diluted in 2× SSC in a water bath at 37 °C for 5 to 15 min. Next, we washed the slides for 2× 5 min in 2× SSC solution and refixed them for 10 min in 3.7% formaldehyde (diluted in 2× SSC). After

fixation, we washed the slides for 2× 5 min in 2× SSC solution, shortly washed in distilled water and dehydrated in an ethanol series (60%, 80%, and 100%), each step 2 min. We then applied denatured hybridization mixture (20 ml each) to each slide, covered with coverslip, placed on a hot plate at 77 °C for 2 min and transferred at 37 °C overnight. We washed slides for 5 min in 2× SSC, transferred for 20 min at 57 °C in 2× SSC, washed 5 min in 2× SSC at RT again, and dehydrated in an ethanol series (60%, 80%, and 100%). Finally, we mounted the slides in VectaShield (Vector, H-1500) supplemented with DAPI (2-(4-aminophenyl)-1H-indole-6-carboxamide). We captured chromosomes under an Olympus AX70 fluorescence microscope equipped with a CCD camera and Imaris software. We used the software GIMP-2.10 and Affinity Photo 2 to process all channels.

Supplementary Material

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

Acknowledgments

We thank Meng Yuan, Bill Cole, and Thomas Gludovacz for the help with plant growth and maintenance, and Alex Harkess and Sarah Carey for the discussion and advice on sex chromosome assembly methods. This research was funded by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada to S.C.H.B. and S.I.W.

Data Availability

Final assemblies are uploaded onto COGE (<https://genomeevolution.org/coge/>) under genome IDs 65183 and 65184 and GenBank under BioProjects PRJNA1046578 (haplotype 1/haplotype A) and PRJNA1046577 (haplotype 2/haplotype B) for *R. hastatulus* and PRJNA1070717 for *R. salicifolius*. All raw reads for the *R. hastatulus* genome assembly are uploaded to the SRA under BioProject accession PRJNA1069061 and raw reads for *R. salicifolius* under BioProject PRJNA1070717. All custom R Scripts are available on GitHub (<https://github.com/SIWLab/XYmaleGenome>).

References

- Abbott JK, Nordén AK, Hansson B. Sex chromosome evolution: historical insights and future perspectives. *Proc Roy Soc B*. 2017;**284**(1854):20162806. <https://doi.org/10.1098/rspb.2016.2806>.
- Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, Imakaev M, Venev SV. Pairtools: from sequencing data to chromosome contacts. *bioRxiv*, <https://doi.org/10.1101/2023.02.13.528389>, 15 February 2023, preprint: not peer reviewed.
- Akagi T, Fujita N, Masuda K, Shirasawa K, Nagaki K, Horiuchi A, Kuwada E, Kunou R, Nakamura K, Ikeda Y, et al. Rapid and dynamic evolution of a giant Y chromosome in *Silene latifolia*. *bioRxiv*, <https://doi.org/10.1101/2023.09.21.558759>, 22 September 2023, preprint: not peer reviewed.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;**215**(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, Mozes E, Strobe PK, Sylla PM, Wagner L, et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol*. 2024;**25**(1):60. <https://doi.org/10.1186/s13059-024-03198-7>.
- Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet*. 2013;**14**(2):113–124. <https://doi.org/10.1038/nrg3366>.
- Bachtrog D. The Y chromosome as a battleground for intragenomic conflict. *Trends Genet*. 2020;**36**(7):510–522. <https://doi.org/10.1016/j.tig.2020.04.008>.
- Báčovský V, Čegan R, Šimoníková D, Hřibová E, Hobza R. The formation of sex chromosomes in *Silene latifolia* and *S. dioica* was accompanied by multiple chromosomal rearrangements. *Front Plant Sci*. 2020;**11**:205. <https://doi.org/10.3389/fpls.2020.00205>.
- Beaudry FEG, Barrett SCH, Wright SI. Genomic loss and silencing on the Y chromosomes of *Rumex*. *Genome Biol Evol*. 2017;**9**(12):3345–3355. <https://doi.org/10.1093/gbe/evx254>.
- Beaudry FEG, Barrett SCH, Wright SI. Ancestral and neo-sex chromosomes contribute to population divergence in a dioecious plant. *Evolution*. 2020;**74**(2):256–269. <https://doi.org/10.1111/evo.13892>.
- Beaudry FEG, Rifkin JL, Peake AL, Kim D, Jarvis-Cross M, Barrett SCH, Wright SI. Effects of the neo-X chromosome on genomic signatures of hybridization in *Rumex hastatulus*. *Mol Ecol*. 2022;**31**(13):3708–3721. <https://doi.org/10.1111/mec.16496>.
- Bergero R, Qiu S, Charlesworth D. Gene loss from a plant sex chromosome system. *Curr Biol*. 2015;**25**(9):1234–1240. <https://doi.org/10.1016/j.cub.2015.03.015>.
- Bieker VC, Battlay P, Petersen B, Sun X, Wilson J, Brealey JC, Bretagnolle F, Nurkowski K, Lee C, Barreiro FS, et al. Uncovering the genomic basis of an extraordinary plant invasion. *Sci Adv*. 2022;**8**(34):eabo5115. <https://doi.org/10.1126/sciadv.abo5115>.
- Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;**18**(4):366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. 2008;**18**(1):188–196. <https://doi.org/10.1101/gr.6743907>.
- Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)*. 2005;**95**(2):118–128. <https://doi.org/10.1038/sj.hdy.6800697>.
- 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**(2):170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Crowson D, Barrett SCH, Wright SI. Purifying and positive selection influence patterns of gene loss and gene expression in the evolution of a plant sex chromosome system. *Mol Biol Evol*. 2017;**34**(5):1140–1154. <https://doi.org/10.1093/molbev/msx064>.
- Cuñado N, Navajas-Pérez R, de la Herrán R, Ruiz Rejón C, Ruiz Rejón M, Santos JL, Garrido-Ramos MA. The evolution of sex chromosomes in the genus *Rumex* (Polygonaceae): identification of a new species with heteromorphic sex chromosomes. *Chromosome Res*. 2007;**15**(7):825–833. <https://doi.org/10.1007/s10577-007-1166-6>.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;**29**(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016;**3**(1):99–101. <https://doi.org/10.1016/j.cels.2015.07.012>.
- Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicebox provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Syst*. 2016;**3**(1):95–98. <https://doi.org/10.1016/j.cels.2016.07.002>.

- Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*. 2008;**9**(1):18. <https://doi.org/10.1186/1471-2105-9-18>.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;**20**(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Fuller ZL, Koury SA, Phadnis N, Schaeffer SW. How chromosomal rearrangements shape adaptation and speciation: case studies in *Drosophila pseudoobscura* and its sibling species *Drosophila persimilis*. *Mol Ecol*. 2019;**28**(6):1283–1301. <https://doi.org/10.1111/mec.14923>.
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012. arXiv:1207.3907, preprint: not peer reviewed.
- Grabowska-Joachimiak A, Kula A, Książczyk T, Chojnicka J, Sliwiska E, Joachimiak AJ. Chromosome landmarks and autosome-sex chromosome translocations in *Rumex hastatulus*, a plant with XX/XY1Y2 sex chromosome system. *Chromosome Res*. 2015;**23**(2):187–197. <https://doi.org/10.1007/s10577-014-9446-4>.
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 2020;**36**(9):2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>.
- Haug-Baltzell A, Stephens SA, Davey S, Scheidegger CE, Lyons E. SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics*. 2017;**33**(14):2197–2198. <https://doi.org/10.1093/bioinformatics/btx144>.
- Hibbins MS, Rifkin JL, Choudhury BI, Voznesenka O, Sacchi B, Yuan M, Gong Y, Barrett SCH, Wright SI. Phylogenomics resolves key relationships in *Rumex* and uncovers a dynamic history of independently evolving sex chromosomes. *bioRxiv*, <https://doi.org/10.1101/2023.12.13.571571>, 14 December 2023, preprint: not peer reviewed.
- Hoagland DR, Snyder WC. Nutrition of strawberry plant under controlled conditions: (a) Effects of deficiencies of boron and certain other elements, (b) susceptibility to injury from sodium salts. *J Am Soc Hortic Sci*. 1933;**30**:288–294.
- Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proc Natl Acad Sci U S A*. 2014;**111**(21):7713–7718. <https://doi.org/10.1073/pnas.1319227111>.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;**30**(9):1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Karafiátová M, Bartoš J, Doležel J. Localization of low-copy DNA sequences on mitotic chromosomes by FISH. *Methods Mol Biol*. 2016;**1429**:49–64. https://doi.org/10.1007/978-1-4939-3622-9_5.
- Kasjaniuk M, Grabowska-Joachimiak A, Joachimiak AJ. Testing the translocation hypothesis and Haldane's rule in *Rumex hastatulus*. *Protoplasma*. 2019;**256**(1):237–247. <https://doi.org/10.1007/s00709-018-1295-0>.
- Kent TV, Uzunović J, Wright SI. Coevolution between transposable elements and recombination. *Philos Trans R Soc B Biol Sci*. 2017;**372**(1736):20160458. <https://doi.org/10.1098/rstb.2016.0458>.
- Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. *Genetics*. 2006;**173**(1):419–434. <https://doi.org/10.1534/genetics.105.047985>.
- Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. *F1000Res*. 2017;**6**:1287. <https://doi.org/10.12688/f1000research.12232.1>.
- Lenormand T, Fyon F, Sun E, Roze D. Sex chromosome degeneration by regulatory evolution. *Curr Biol*. 2020;**30**(15):3001–3006.e5. <https://doi.org/10.1016/j.cub.2020.05.052>.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;**34**(18):3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;**25**(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Löve Á. Chromosome number reports XCII. *Taxon*. 1986;**35**(3):610–613. <https://doi.org/10.1002/j.1996-8175.1986.tb00821.x>.
- Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkness A, Emms D, Goodstein DM, Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*. 2022;**11**:e78526. <https://doi.org/10.7554/eLife.78526>.
- Lowry DB, Willis JH. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol*. 2010;**8**(9):e1000500. <https://doi.org/10.1371/journal.pbio.1000500>.
- Mandáková T, Pouch M, Brock JR, Al-Shehbaz IA, Lysak MA. Origin and evolution of diploid and allopolyploid *Camelina* genomes were accompanied by chromosome shattering. *Plant Cell*. 2019;**31**(11):2596–2612. <https://doi.org/10.1105/tpc.19.00366>.
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;**38**(10):4647–4654. <https://doi.org/10.1093/molbev/msab199>.
- Moraga C, Branco C, Rougemont Q, Veltsos P, Jedlicka P, Muyle A, Hanique M, Tannier E, Liu X, Mendoza-Galindo E, et al. The *Silene latifolia* genome and its giant Y chromosome. *bioRxiv*, <https://doi.org/10.1101/2023.09.21.558754>, 22 September 2023, preprint: not peer reviewed.
- Mrnjavac A, Khudiakova KA, Barton NH, Vicoso B. Slower-X: reduced efficiency of selection in the early stages of X chromosome evolution. *Evol Lett*. 2023;**7**(1):4–12. <https://doi.org/10.1093/evlett/grac004>.
- Muyle AM, Seymour DK, Lv Y, Huettel B, Gaut BS. Gene body methylation in plants: mechanisms, functions, and important implications for understanding evolutionary processes. *Genome Biol Evol*. 2022;**14**(4):evac038. <https://doi.org/10.1093/gbe/evac038>.
- Novák P, Ávila Robledillo L, Kobličková A, Vrbová I, Neumann P, Macas J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res*. 2017;**45**(12):e111. <https://doi.org/10.1093/nar/gkx257>.
- Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc*. 2020;**15**(11):3745–3776. <https://doi.org/10.1038/s41596-020-0400-y>.
- Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*. 2013;**29**(6):792–793. <https://doi.org/10.1093/bioinformatics/btt054>.
- Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;**32**(2):292–294. <https://doi.org/10.1093/bioinformatics/btv566>.
- Orr HA, Kim Y. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics*. 1998;**150**(4):1693–1698. <https://doi.org/10.1093/genetics/150.4.1693>.
- Ou S, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol*. 2018;**176**(2):1410–1422. <https://doi.org/10.1104/pp.17.01310>.
- Ou S, Jiang N. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob DNA*. 2019;**10**(1):48. <https://doi.org/10.1186/s13100-019-0193-0>.
- Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellings AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol*. 2019;**20**(1):275. <https://doi.org/10.1186/s13059-019-1905-y>.
- Papadopoulos AST, Chester M, Ridout K, Filatov DA. Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc Natl Acad Sci U S A*. 2015;**112**(42):13021–13026. <https://doi.org/10.1073/pnas.1508454112>.

- Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, *et al.* Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* 2020;**21**(1):177. <https://doi.org/10.1186/s13059-020-02097-x>.
- Pickup M, Barrett SCH. The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. *Ecol Evol.* 2013;**3**(3):629–639. <https://doi.org/10.1002/ece3.465>.
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 2016;**26**(3):342–350. <https://doi.org/10.1101/gr.193474.115>.
- R Core Team. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing; 2022.
- Rice WR. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution.* 1987;**41**(4):911–914. <https://doi.org/10.2307/2408899>.
- Rifkin JL, Beaudry FEG, Humphries Z, Choudhury BI, Barrett SCH, Wright SI. Widespread recombination suppression facilitates plant sex chromosome evolution. *Mol Biol Evol.* 2021;**38**(3):1018–1030. <https://doi.org/10.1093/molbev/msaa271>.
- Rifkin JL, Hnatovska S, Yuan M, Sacchi BM, Choudhury BI, Gong Y, Rastas P, Barrett SCH, Wright SI. Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. *Philos Trans R Soc B Biol Sci.* 2022;**377**(1850):20210226. <https://doi.org/10.1098/rstb.2021.0226>.
- Sandler G, Beaudry FEG, Barrett SCH, Wright SI. The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants. *Evol Lett.* 2018;**2**(4):368–377. <https://doi.org/10.1002/evl3.60>.
- Schubert V, Ruban A, Houben A. Chromatin ring formation at plant centromeres. *Front Plant Sci.* 2016;**7**:28. <https://doi.org/10.3389/fpls.2016.00028>.
- Shi J, Liang C. Generic repeat finder: a high-sensitivity tool for genome-wide *de novo* repeat detection. *Plant Physiol.* 2019;**180**(4):1803–1815. <https://doi.org/10.1104/pp.19.00386>.
- Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics.* 2021;**37**:1639–1643. <https://doi.org/10.1093/bioinformatics/btaa1016>.
- Smit, AFA, Hubley, R. 2008. *Repeat modeler open-1.0*. <http://www.repeatmasker.org>.
- Smith BW. The evolving karyotype of *Rumex hastatulus*. *Evolution.* 1964;**18**(1):93–104. <https://doi.org/10.2307/2406423>.
- Song B, Marco-Sola S, Moreto M, Johnson L, Buckler ES, Stitzer MC. AnchorWave: sensitive alignment of genomes with high sequence diversity, extensive structural polymorphism, and whole-genome duplication. *Proc Natl Acad Sci U S A.* 2022;**119**(1):e2113075119. <https://doi.org/10.1073/pnas.2113075119>.
- Su W, Gu X, Peterson T. TIR-Learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. *Mol Plant.* 2019;**12**(3):447–460. <https://doi.org/10.1016/j.molp.2019.02.008>.
- Subrini J, Turner J. Y chromosome functions in mammalian spermatogenesis. *eLife.* 2021;**10**:e67345. <https://doi.org/10.7554/eLife.67345>.
- Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, Huang K, Ostevik KL, Drummond EBM, Imerovski I, *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature.* 2020;**584**(7822):602–607. <https://doi.org/10.1038/s41586-020-2467-6>.
- Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H, *et al.* MCScanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;**40**(7):e49. <https://doi.org/10.1093/nar/gkr1293>.
- Wei KH-C, Gibilisco L, Bachtrog D. Epigenetic conflict on a degenerating Y chromosome increases mutational burden in *Drosophila* males. *Nat Commun.* 2020;**11**(1):5537. <https://doi.org/10.1038/s41467-020-19134-9>.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al.* A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 2007;**8**(12):973–982. <https://doi.org/10.1038/nrg2165>.
- Xiong W, He L, Lai J, Dooner HK, Du C. HelitronScanner uncovers a large overlooked cache of *Helitron* transposons in many plant genomes. *Proc Natl Acad Sci U S A.* 2014;**111**(28):10263–10268. <https://doi.org/10.1073/pnas.1410068111>.
- Zhang L, Li X, Ma B, Gao Q, Du H, Han Y, Li Y, Cao Y, Qi M, Zhu Y, *et al.* The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Mol Plant.* 2017;**10**(9):1224–1237. <https://doi.org/10.1016/j.molp.2017.08.013>.
- Zhang R-G, Li G-Y, Wang X-L, Dainat J, Wang Z-X, Ou S, Ma Y. TESorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. *Hortic Res.* 2022;**9**:uhac017. <https://doi.org/10.1093/hr/uhac017>.
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 2023;**39**(1):btac808. <https://doi.org/10.1093/bioinformatics/btac808>.