Title

Enrichment of hard sweeps on the X chromosome compared to autosomes in six Drosophila species.

Journal

Genetics, 226(4)

Authors

Harris, Mariana
Kim, Bernard
Garud, Nandita

Publication Date

2024-04-03

DOI

10.1093/genetics/iyae019

Peer reviewed

OXFORD GENETICS

GENETICS, 2024, 226(4), iyae019
https://doi.org/10.1093/genetics/iyae019
Advance Access Publication Date: 16 February 2024
Investigation

Enrichment of hard sweeps on the X chromosome compared to autosomes in six *Drosophila* species

Mariana Harris [1], Bernard Y. Kim,[2] Nandita Garud[3,4,*]

[1]Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA
[2]Department of Biology, Stanford University, Stanford, CA 94305, USA
[3]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095, USA
[4]Department of Human Genetics, University of California Los Angeles, Los Angeles, CA 90095, USA

*Corresponding author: Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095, USA; Department of Human Genetics, University of California Los Angeles, Los Angeles, CA 90095, USA. Email: ngarud@ucla.edu

The X chromosome, being hemizygous in males, is exposed one-third of the time increasing the visibility of new mutations to natural selection, potentially leading to different evolutionary dynamics than autosomes. Recently, we found an enrichment of hard selective sweeps over soft selective sweeps on the X chromosome relative to the autosomes in a North American population of *Drosophila melanogaster*. To understand whether this enrichment is a universal feature of evolution on the X chromosome, we analyze diversity patterns across 6 commonly studied *Drosophila* species. We find an increased proportion of regions with steep reductions in diversity and elevated homozygosity on the X chromosome compared to autosomes. To assess if these signatures are consistent with positive selection, we simulate a wide variety of evolutionary scenarios spanning variations in demography, mutation rate, recombination rate, background selection, hard sweeps, and soft sweeps and find that the diversity patterns observed on the X are most consistent with hard sweeps. Our findings highlight the importance of sex chromosomes in driving evolutionary processes and suggest that hard sweeps have played a significant role in shaping diversity patterns on the X chromosome across multiple *Drosophila* species.

Keywords: hard sweeps; X chromosome; adaptation; *Drosophila*

Introduction

The X chromosome has long been a subject of substantial interest in evolutionary biology due to its unique features that set it apart from the autosomes. Notably, the X harbors genes responsible for speciation, fertility, sexual dimorphism, and brain function (Rice 1984; Turelli and Orr 1995; Coyne and Orr 1998; Saifi and Chandra 1999; Skuse 2005; Payseur *et al.* 2018), highlighting its biological importance. Moreover, previous work suggests that the X chromosome may serve as a potential target of sexually antagonistic selection (Dean and Mank 2014; Patten 2019; Glaser-Schmitt *et al.* 2021), further emphasizing its significance in evolution. Consequently, studying adaptation on the X chromosome, and how it differs from that of autosomes, can provide insights into the mechanisms driving genetic diversity, sexual selection, and speciation, thereby deepening our understanding of the broader processes that shape genetic variation across populations.

Adaptation on the X may differ from that of the autosomes due to 2 key differences. First, the X is expected to have a lower effective population size (N_{eX}) compared to autosomes (N_{eA}), leading to a decreased influx of new mutations. Second, due to male hemizygosity, new mutations on the X of males are immediately exposed to natural selection. This increased exposure to selection may lead to a higher probability of fixation of new recessive beneficial mutations ("Faster-X" effect; Charlesworth *et al.* 1987) and a more efficient purging of deleterious variation on the X compared to autosomes, leading to lower levels of standing genetic variation on the X. Thus, as a consequence of these 2 factors, at the onset of positive selection, there will be a lower adaptive mutational supply on the X, resulting in more gradual rates of adaptation, or in other words, fewer haplotypes rising to high frequency bearing the adaptive allele (Orr and Betancourt 2001; Vicoso and Charlesworth 2006, 2009; Charlesworth *et al.* 2018).

Adaptation leaves behind distinct signatures in the genome. The classic signature, referred to as a hard selective sweep, occurs when a single adaptive mutation rises in frequency, resulting in deep dips in diversity in the vicinity of the adaptive locus (Smith and Haigh 1974; Kaplan *et al.* 1989). By contrast, a different signature known as a soft selective sweep occurs when multiple adaptive mutations on distinct haplotypes sweep through the population simultaneously, not necessarily causing dips in diversity (Hermisson and Pennings 2005, 2017; Pennings and Hermisson 2006a; Messer and Petrov 2013). Recently, we found evidence of an enrichment of hard sweeps on the X chromosome compared to autosomes in a North American population of *D. melanogaster* (Harris and Garud 2023), suggesting that the X chromosome is subject to different evolutionary dynamics than the autosomes. Whether an enrichment of hard sweeps on the X is a universal feature of molecular evolution has yet to be determined, as only a few species have been shown to have a higher prevalence of hard sweeps on the X compared to the autosomes (Nam *et al.* 2015; Harris and Garud 2023). Quantifying the prevalence of hard vs soft sweeps in natural populations has been of

Received on 06 December 2023; accepted on 18 January 2024

great interest and debate (Peter *et al.* 2012; Assaf *et al.* 2015; Schrider *et al.* 2015; Schrider and Kern 2016; Harris, Sackman, and Jensen 2018; Feder *et al.* 2021; Garud *et al.* 2021). Thus, understanding the prevalence of hard and soft sweeps more broadly is crucial as it can shed light on common mechanisms that underlie adaptation in natural populations.

To understand if the X is generically enriched for hard sweeps in many species, we analyze population genomic data from 6 *Drosophila* species. By leveraging whole-genome samples from multiple species and populations (Arbiza *et al.* 2014; Nam *et al.* 2015; McGrath 2022), we can identify trends that are the norm across species, as well as exceptions that are indicative of the unique biology of individual species. However, four of the populations we analyze in this study have small sample sizes ($n = 7$–23 samples), making it difficult to conduct the same haplotype-based scan used in our previous work (Garud *et al.* 2015; Harris and Garud 2023). Additionally, each of these species has a unique demographic history for which we do not have accurate models, rendering our previous simulation-based approach for classifying putative sweeps as hard and soft challenging (Harris and Garud 2023). To overcome these challenges, we use a combination of single-nucleotide diversity and haplotype homozygosity statistics to analyze the patterns of diversity on the autosomes and the X chromosome across species, specifically looking for evidence of hard sweeps on the X that is inconsistent with other neutral and selective forces including demography, background selection (BGS), and soft selective sweeps. Our empirical and simulation analyses show evidence that hard sweeps have played a significant role in shaping diversity patterns on the X chromosome in multiple *Drosophila* species suggesting that hard sweeps on the X are the norm rather than the exception.

Methods

Data

We analyzed data from 6 *Drosophila* species from 7 populations [*D. melanogaster* from Zambia (ZI), *D. melanogaster* from Raleigh (RA), *Drosophila simulans*, *Drosophila sechellia*, *Drosophila mauritiana*, *Drosophila santomea*, and *Drosophila teissieri*], which we downloaded and processed as follows:

For both *D. melanogaster* populations, data are publicly available as part of the Drosophila Genome Nexus data set (Lack *et al.* 2015) and can be downloaded from www.johnpool.net. We downloaded 205 Drosophila Genetic Reference Panel (DGRP) genomes from RA, North Carolina and 197 DPGP3 genomes from ZI. For our analysis, we used 100 genomes from each population, previously processed (Harris and Garud 2023) to remove individuals having high IBD with one another, as well as residual heterozygosity, and high levels of missing data.

For *D. simulans*, we downloaded data from 170 inbred lines from a North American population (Signor *et al.* 2018) available at https://zenodo.org/record/154261#.YzMzty2z3jC. For the remaining 4 species, we obtained genomes from NCBI's RefSeq and short-read data from NCBI's Short Read Archive (Garrigan *et al.* 2014; Turissini and Matute 2017; Meany *et al.* 2019; Serrato-Capuchina *et al.* 2021) (Supplementary Table 1).

For *D. mauritiana*, *D. sechellia*, *D. teissieri*, and *D. santomea*, variant calling was performed with the NVIDIA Clara Parabricks pipeline v4.0.0 (https://docs.nvidia.com/clara/parabricks/4.0.0/index.html), a reimplementation of variant calling tools including BWA-mem (Li 2013) and GATK4 (Van der Auwera and O'Connor 2020) optimized for running on NVIDIA graphics processing units. Following the GATK best practices for germline variant calling (https://gatk.broadinstitute.org/hc/en-us/sections/360007226651), we mapped each sample's reads to the appropriate reference genome, sorted and removed PCR duplicates, generated single-sample Genomic Variant Call Format (GVCF)s, and then performed joint genotyping with HaplotypeCaller. From this initial set of variant calls, we removed sites with quality scores QUAL < 30.0 to obtain a bootstrap set of high-confidence calls for another round of variant calling (https://gatk.broadinstitute.org/hc/en-us/articles/360035890531-Base-Quality-Score-Recalibration-BQSR-). We performed base quality score recalibration on the mapped reads using the bootstrap set then once again generated single sample GVCFs that were used for joint genotyping with HaplotypeCaller to ultimately generate a VCF file for each species.

To remove any low-quality variant sites, we applied GATK recommended hard filters (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants), summarized as follows: FS > 60, QD < 2.0, MQ < 4, MQRankSum < -12.5, QUAL < 30.0, SOR > 3.0, and ReadPosRankSum < -8.0. Additionally, we excluded sites that were not uniquely mappable (Supplementary Fig. 1). We also filtered out sites lying within repetitive elements as predicted by RepeatMasker (Smit et al. 2013-2015). Only biallelic SNPs were considered in our analysis.

Next, we excluded invariant sites of poor quality. To do so, we filtered entire regions, which included both variant and invariant sites, with poor depth and quality statistics. We used banded GVCF files, where sites of similar quality get concatenated into a band. We excluded intervals that failed to meet the depth and quality criteria: ('MIN(FMT/DP) > 10 & MIN(FMT/GQ) > 30'). Next, we combined the results from the GATK hard filters applied to variant sites with the resulting regions—comprising both variant and invariant sites—that passed the filter applied to the banded GVFC files. This gave us the total number of callable sites in the genome. When computing statistics using a sliding window approach (see section below), we excluded windows that overlapped 50% or more with low-quality intervals.

High-density repeat regions are often associated with centromeric regions, which consist of highly homogeneous tandem repeats and are known to experience low rates of recombination (Mather 1939; Levine 1955; Vincenten *et al.* 2015). Consequently, these regions tend to exhibit reduced diversity and increased homozygosity, which can lead to false positive signals of selection. To mitigate potential confounding effects of repeats on selection inferences, we calculated the proportion of base pairs identified as repeats within 50-kb windows. In addition to removing individual sites masked as repeats, we removed entire windows in which 20% or more of the sites were marked as repeats by RepeatMasker.

Diversity statistics and haplotype homozygosity measured from data

We annotated which SNPs lie in exon, intron, or intergenic regions using RefSeq (O'Leary *et al.* 2016) annotations. The corresponding assembly accession numbers for each species are as follows: GCF_004382145.1 (*D. mauritiana*), GCF_004382195.2 (*D. sechellia*), GCF_016746235.2 (*D. teissieri*), GCF_016746245.2 (*D. santomea*), and GCF_000001215.2 (*D. melanogaster*). For *D. simulans*, we used the reference genome and corresponding gff file provided by Rebekah Rogers and Peter Andolfatto (Rogers *et al.* 2014). We used BEDtools v.2.3.0 (Quinlan and Hall 2010) to separate the data into exon, introns, or intergenic regions. If any position could be included in multiple categories, we annotated the position prioritizing exons followed by introns and finally intergenic regions. Additionally, we excluded regions less than 40 bp long.

To understand the influence of selection across genomic regions on the X chromosome vs autosomes, we calculated nucleotide diversity in autosomes (π_A) and the X chromosome (π_X) in each of the exons, introns, and intergenic regions. We next computed the ratio of X to autosomal diversity (π_X/π_A) by running 1,000 bootstrap replicates in which we sampled exons, introns, and intergenic regions at random, computing the mean π_X, mean π_A, and mean(π_X)/mean(π_A) in each sample. We approximated π by $\pi \asymp 2pq$, with p the frequency of the major allele at a given site and $q = 1 - p$.

To identify regions of the genome that may have exceptionally low diversity due to selection, we also computed π/bp in nonoverlapping windows of 10, 20, and 50 kb (Supplementary Fig. 2) across the autosomes and the X chromosome of all species. We then defined a low diversity threshold as X% of the π/bp average of each chromosome with X = 20 or 30%. We labeled windows with π/bp below this threshold as low diversity windows, further investigated as putative sweep regions. For the main analysis of this work, 20-kb windows and a threshold of 20% of the average π/bp were used. To test for differences in quality in windows below and above the defined threshold, we computed the proportion of missing data and quality score per site and found similar distributions across these categories (Supplementary Fig. 3).

Additionally, we computed haplotype or multilocus genotype homozygosity in windows of 20 SNPs long on the X chromosome, with the expectation that haplotype homozygosity should be elevated in selective sweeps but not BGS (Wall and Pritchard 2003; Enard *et al.* 2014; Garud *et al.* 2015; Schrider 2020). In species for which we had phased data (*D. melanogaster*, *D. simulans*, and *D. mauritiana*), we computed the expected haplotype homozygosity (Garud *et al.* 2015) defined as $H = \sum_{i=1}^{n} p_i^2$, where p_i is the frequency of the ith most common haplotype in a sample with n distinct haplotypes. For the remaining species, we computed the expected multilocus genotype homozygosity (Harris, Garud, and DeGiorgio 2018) defined as $G = \sum_{i=1}^{n} q_i^2$, where q_i is the frequency of the ith most common multilocus genotype in a sample with n distinct haplotypes. In contrast to a phased haplotype, where the allelic state for each site is known, a multilocus genotype is a string that represents the diploid state of the individual where each site is labeled as either homozygous for the reference allele, homozygous for the alternate allele, or heterozygous. A high recombination rate is expected to break haplotypes and thus decrease both H and G, whereas with a low recombination rate, haplotypes may drift to high frequency and thus increase H, and by extension G.

In addition to measuring H and G in 20 SNP windows, we tested windows of length 50 and 10 SNPs long. However, due to the small sample size of most species, the probability of observing the same multilocus genotype twice becomes small with longer window sizes, making it difficult to capture any signal in the 50 SNP window case. With 10 SNP windows, we also found an elevation of homozygosity in low diversity windows. However, we opted for 20 SNP windows as smaller window sizes may increase homozygosity due to genetic drift.

Simulation analysis

To understand the evolutionary processes responsible for the patterns of π/bp and haplotype homozygosity observed in the data, we simulated a variety of evolutionary models for the X chromosome and autosomes using SLiM 3.7 (Haller and Messer 2019). The scenarios simulated included neutrality, sex bias, low recombination rate, mutation rate bias, bottlenecks, BGS, hard sweeps, and soft sweeps as described below.

For all models with a fixed population size (excluding bottleneck models), we simulate a constant $N_e = 10^6$ population. An N_e of this order of magnitude is reasonable for most of the species analyzed given mean nucleotide diversity levels, with exception of *D. sechellia*, which has been shown to have a lower $N_e \sim 10^5$ (Legrand *et al.* 2009). SLiM is a forward-in-time simulator, which makes simulating populations with $N_e > 5 \times 10^5$ unfeasible due to memory requirements. As in our previous work (Harris and Garud 2023), we rescaled our simulations using a constant factor of Q = 50.

In all our simulations, we modeled a 20-kb region with a recombination rate of $r = 5 \times 10^{-7}$ cM/bp (unless otherwise specified) and a neutral mutation rate of $\mu = 1 \times 10^{-9}$, both rescaled by Q = 50. For the simulations that include selection, we assumed that mutations on the X of males experience the same fitness effect as that of a homozygous female (i.e. dosage compensation). We ran a total of 100 simulations for each model and ran a $10N_e$ burn-in for every simulation.

Neutral models

We first simulated a completely neutral model with a female-to-male ratio of 1 and equal mutation rates between the sexes. Next, we introduced scenarios that can differentially impact X-linked and autosomal diversity. We simulated female and male sex bias varying the sex ratio as 2:1, 5:1, and 7:1 for each scenario, as well as a model with a lower X-linked mutation rate reducing the X-linked mutation rate such that the ratio μ_X/μ_A was equal to 0.1, 0.5, 0.75, or 0.9. We also tested the effect of regions of low recombination by running simulations with recombination reduced to 1, 10, 20, and 50% of the original r.

The expected lower population size of the X can result in stronger drift that can be exacerbated in a bottleneck. For this reason, we considered 2 bottleneck models that were fit to π/bp and S/bp in short introns from DGRP data set (Garud *et al.* 2015, 2021): (1) a severe and short bottleneck with a bottleneck N_e of $0.002N_{e,\ \text{ancestral}}$ for $0.0002 * 2N_{e,\ \text{ancestral}}$ generations and (2) a shallow and long bottleneck with a bottleneck N_e of $0.4N_{e,\ \text{ancestral}}$ for $0.056 * 2N_{e,\ \text{ancestral}}$ generations.

Background selection

BGS can result in stronger dips in diversity on the X compared to autosomes (Charlesworth *et al.* 1993, 1995; Stephan 2010). To consider this scenario, we simulated a constant $N_e = 10^6$ model with deleterious variation. The selection coefficients for deleterious mutations were gamma distributed with mean and shape parameter of −0.000133 and 0.35, respectively (Huber *et al.* 2017). We varied the percentage of deleterious mutations such that 10, 50, or 80% of incoming mutations were deleterious. Additionally, we simulated BGS and varied the X-linked mutation rate in both bottleneck models described previously to test whether the effect of both processes can produce the patterns in the data. We did this for $r = 2.5e^{-7}$ and $r = 5e^{-7}$ cM/bp. Furthermore, we simulated 1-Mb chromosomes varying the recombination rate ($r = 0$, $1e^{-8}$, and $5e^{-7}$ cM/bp) to assess the effect of a higher mutation load and recombination rate in the context of BGS.

Selective sweeps

To understand the effects of positive selection on patterns of diversity on the X chromosome and autosomes, we simulated hard and soft sweeps. To model hard sweeps, we introduced a single adaptive mutation to the center of the haplotype ($\theta_a = 0.01$) and restarted the simulation if the adaptive mutation was lost. We conditioned on fixation of the sweep and sampled 100 haplotypes.

To simulate soft sweeps, we introduced adaptive mutations recurrently to the center of the haplotype at a rate defined by $\theta_a = 0.1$, 1, and 10, where $\theta_a = 4N_e\mu_a$ with μ_a the adaptive mutation rate. We conditioned our simulations on the fixation of the sweep, after which we took a sample of $n = 100$ haplotypes. We verified that our simulation represented a soft sweep by only including samples with 2 or more mutational origins in our analysis. For both hard and soft sweeps, we varied the strength of selection such that $N_e s_b = 20$, 200, or 2,000.

Identification of shared genes under selection in multiple species

To investigate whether similar functions are selected for on the X in multiple species, we obtained the genes intersecting windows below and above the low diversity threshold using BEDtools v.2.3.0 and the annotation files from the reference genome for each species in our study. Next, we used the gene IDs and matched them to their corresponding ortholog IDs in orthoDB v11 (Kuznetsov *et al.* 2023). We then looked for overlapping orthologs across species and obtained the proportion of orthologs that overlap over 2, 3, or 4 species in low vs high diversity windows.

To understand whether there were more shared regions under selection in low vs high diversity regions across species, we obtained 100 random samples of n orthologs from high diversity regions with n the number of orthologs in low diversity regions per species. From this, we computed the mean proportion of shared orthologs under selection as well as the 95% confidence interval in high diversity regions.

Results

We analyzed population genomic data of 6 different *Drosophila* species from 7 populations, including *D. melanogaster* ($n = 100$ ZI, $n = 100$ RA), *D. simulans* ($n = 170$), *D. sechellia* ($n = 23$), *D. mauritiana* ($n = 15$), *D. santomea* ($n = 7$), and *D. teissieri* ($n = 11$) (Supplementary Table 2). To understand not only how selection varies across species but also across populations within a species, we included a derived *D. melanogaster* population from RA as well as a population from ZI that is presumed to be within the ancestral range of the species (Pool *et al.* 2012). To investigate whether hard sweeps are enriched across these species, we first compare the patterns of nucleotide diversity (π) between the autosomes and the X chromosome in the data. Next, we analyze haplotype homozygosity in low diversity regions on the X compared to the rest of the chromosome. Finally, we simulate a variety of models, both without and with selection, to test whether any of these can generate the patterns observed in the data.

Patterns of X vs autosomal diversity across the genomes of 6 *Drosophila* species

When a population is not subject to any selective forces, has the same mutation rate across sexes, a male-to-female ratio of 1, and a constant effective population size (N_e), the diversity on the X chromosome (π_X) is expected to be 3/4 of the autosomal diversity (π_A) (Vicoso and Charlesworth 2006). However, such a simple model rarely captures the complex dynamics of natural populations, rather, it is likely that different evolutionary forces affect the X and autosome differently, leading to deviations from the $\pi_X/\pi_A = 0.75$ expectation. Across the 7 populations studied, we observed

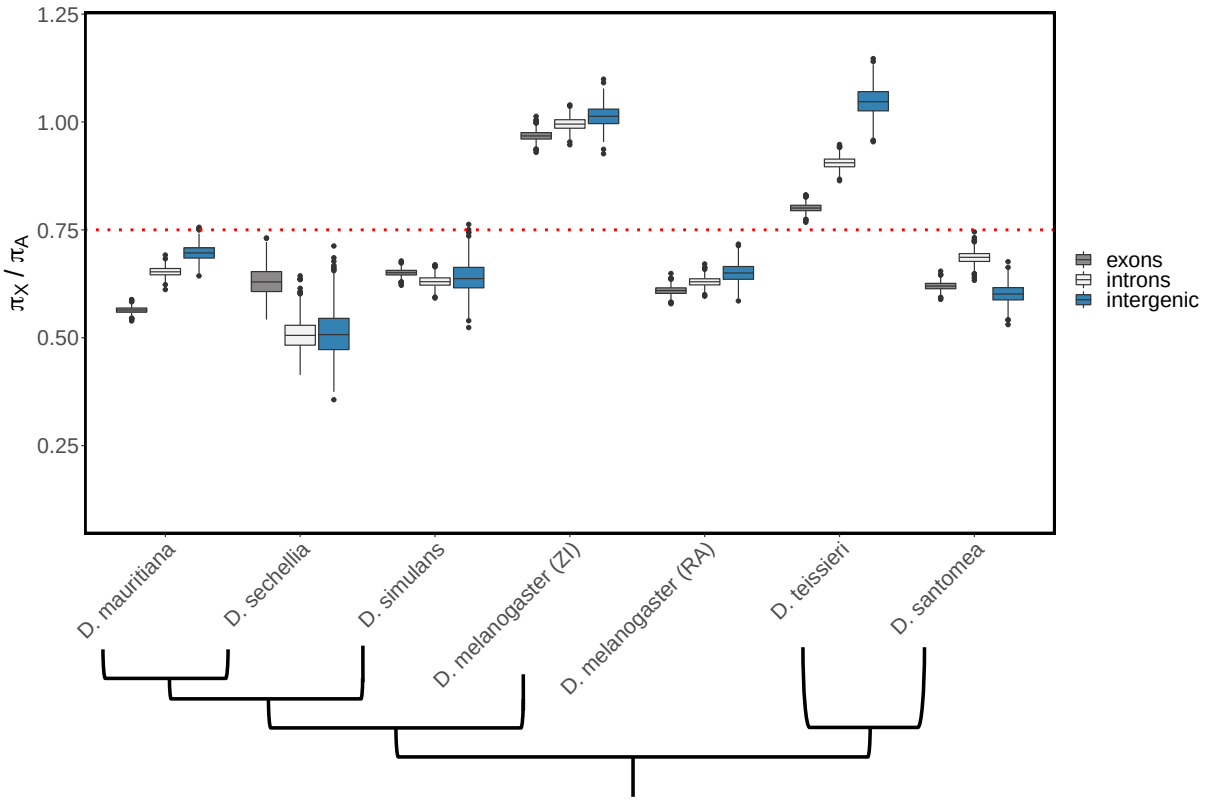

Fig. 1. Nucleotide diversity (π/bp) in exons, introns, and intergenic regions on the autosomes and the X chromosome of 6 *Drosophila* species. π_X/π_A is lower on exons than on introns and intergenic regions in most species. The red dotted line corresponds to the 0.75 expectation under a model with an unbiased sex ratio and equal mutation rates on the X and autosomes. For each category, we performed 1,000 bootstrap replicates (see *Methods*). The tree shown below the X-axis represents the phylogenetic relationship between species as described in FlyBase (Gramates *et al.* 2022).

a significantly lower diversity on the X compared to autosomes with 5 out of 7 populations showing π_X/π_A below 0.75 (Fig. 1). The other 2 populations [*D. melanogaster* (ZI) and *D. teissieri*] also deviated from the $\pi_X/\pi_A = 0.75$ expectation but in the opposite direction, with $\pi_X/\pi_A > 0.75$. This trend has been previously reported for African populations of *D. melanogaster* and has been attributed to other evolutionary forces such as female sex bias (Kauer *et al.* 2002; Singh *et al.* 2007; Pool *et al.* 2012).

Functionally important regions of the genome, such as exons, are expected to be subject to selection, positive or negative (McVicker *et al.* 2009; Arbiza *et al.* 2014; Nam *et al.* 2015). Hence, to investigate the influence of selection across species, we divided the data into 3 regions: exons, introns, and intergenic regions and computed the ratio of X to autosomal diversity (π_X/π_A) for each region. We found a stronger reduction in diversity in exons compared to introns and intergenic regions for most species (Fig. 1), pointing to the functional importance of exons (Nam *et al.* 2015). Moreover, we found that π_X/π_A was consistently below 0.75 for all regions for most species (Fig. 1), suggesting more directional selection on the X compared to autosomes. However, we noted an exception in *D. sechellia*, where exons exhibited higher levels of diversity than introns and intergenic regions. This behavior could be explained by the high levels of introgression from *D. simulans* to *D. sechellia* (Garrigan *et al.* 2012; Matute and Ayroles 2014; Schrider *et al.* 2018). Introgression tends to be higher in nonfunctional regions of the genome and is generally higher on the autosomes than on the X due to the involvement of sex chromosomes in hybrid incompatibilities (Muirhead and Presgraves 2016; Turissini and Matute 2017; Fraïsse and Sachdeva 2021). This could lead to a greater difference between X and autosomal diversity in intergenic regions compared to exons, potentially explaining the patterns observed in *D. sechellia*'s data.

To identify potential regions under selection, we calculated π/bp in 20-kb windows across each chromosome for every species. Next, we labeled the analysis windows as low diversity windows if they fell below a diversity threshold set as 20% of the chromosomal π/bp average. Importantly, we computed this threshold for each chromosome separately to normalize for variance in coalescence time across chromosomes. We found that, for all species, the X chromosome showed a higher proportion of windows below the low diversity threshold compared to the autosomes (Fig. 2; Supplementary Fig. 4). Remarkably, these low diversity windows often exhibited π/bp values significantly lower than the defined threshold with values as low 0.5% of the chromosomal average (Supplementary Table 3; Supplementary Fig. 5). Furthermore, we observed multiple instances of consecutive low diversity windows extending up to 200 kb, shown as the yellow shaded regions of Fig. 2a. To validate our findings, we repeated this analysis using 10- and 50-kb analysis windows as well as 2 different low diversity thresholds. In all cases, we observed an increased proportion of low diversity windows on the X compared to autosomes (Supplementary Fig. 2).

Elevated homozygosity can arise from different processes such as positive selection or population bottlenecks (Wall and Pritchard 2003). Therefore, to gain deeper insights into the potential mechanisms responsible for the dips in diversity on the X, we computed haplotype homozygosity (H) across windows below and above the low diversity threshold (see *Methods*; Fig. 3) in species with phased genomes (*D. melanogaster*, *D. simulans*, and *D. mauritiana*). For species with unphased genomes, we computed multilocus genotype identity (G) (Harris, Garud, and DeGiorgio 2018; see *Methods*). We found that in all populations, windows below the low diversity threshold showed a significantly elevated homozygosity compared to the rest of the chromosome (1-sided Wilcoxon rank-sum test $P < 0.05$; Fig. 3). Below, we examine the role of selection in generating these differential patterns between the X and the autosomes.

Neutral models cannot recapitulate diversity patterns observed in the data

The differing diversity patterns on the X vs autosomes could be the result of distinct population genetic forces including (1) sex-biased demography, (2) differences in mutation rates between males and females, (3) lower recombination rates, (4) stronger drift on the X given its smaller N_e, and/or (5) more efficient natural selection on the X due to male hemizygosity (Betancourt *et al.* 2004; Vicoso and Charlesworth 2009; Arbiza *et al.* 2014; Nam *et al.* 2015). To gain a better understanding on how these different evolutionary processes influence the relative X to autosomal diversity, we quantified π_X, π_A, and haplotype homozygosity across a wide variety of scenarios simulated with the population genetics simulator SLiM (Haller and Messer 2019).

Sex bias

Our simulations show that female bias can increase π_X/π_A above the 0.75 expectation (Fig. 4a), recapitulating the patterns observed in the Zambian population of *D. melanogaster* and consistent with evidence of female bias in African *D. melanogaster* populations previously reported in the literature (Kauer *et al.* 2002; Dieringer *et al.* 2005; Thornton and Andolfatto 2006; Singh *et al.* 2007; Pool *et al.* 2012) (Figs. 1 and 4a). *D. teissieri* also has $\pi_X/\pi_A > 0.75$, which could also be consistent with female bias. However, none of our female bias models produced elevated haplotype homozygosity compared to neutrality (Fig. 4b).

Additionally, we found that male sex bias can decrease π_X such that $\pi_X/\pi_A < 0.75$, which could potentially explain low π_X values observed in some species (Fig. 1). However, only the most extreme case of male bias (1:7 female–male bias) can marginally increase haplotype homozygosity compared to neutrality (Fig. 4b). To the best of our knowledge, such an extreme male–sex bias in the species studied has not been reported in the literature and therefore seems unlikely.

Recombination rate variation

Another possibility is that regions with a low recombination rate produce dips in diversity and high haplotype homozygosity as observed in the data. The footprint of a hard selective sweep is inversely proportional to the recombination rate ($\sim s/r$), due to high recombination rates breaking linkage and impeding the formation of long haplotypes at high frequency (Smith and Haigh 1974; Gillespie 2004). Consequently, it is reasonable to anticipate lower recombination rates in our putative hard sweep regions when compared to the overall genomic landscape. For *D. melanogaster*, where there is a detailed recombination map available, we find that, on average, recombination rates are 37 and 48% lower in the low diversity regions of the RA and ZI populations, respectively (Supplementary Fig. 6; Comeron *et al.* 2012). To understand if a lower recombination rate can explain all the signatures in the data, we simulate windows with reduced recombination rates (r_{low}), such that $r_{low}/r = 0.5$, 0.2, 0.1, and 0.01. We found that while a low recombination rate can elevate haplotype homozygosity, it cannot generate dips of diversity below the low diversity threshold (Fig. 4; Supplementary Fig. 7).

Mutation rates

Lower X-linked mutation rates could result in stronger dips in diversity on the X compared to autosomes. A lower X-linked

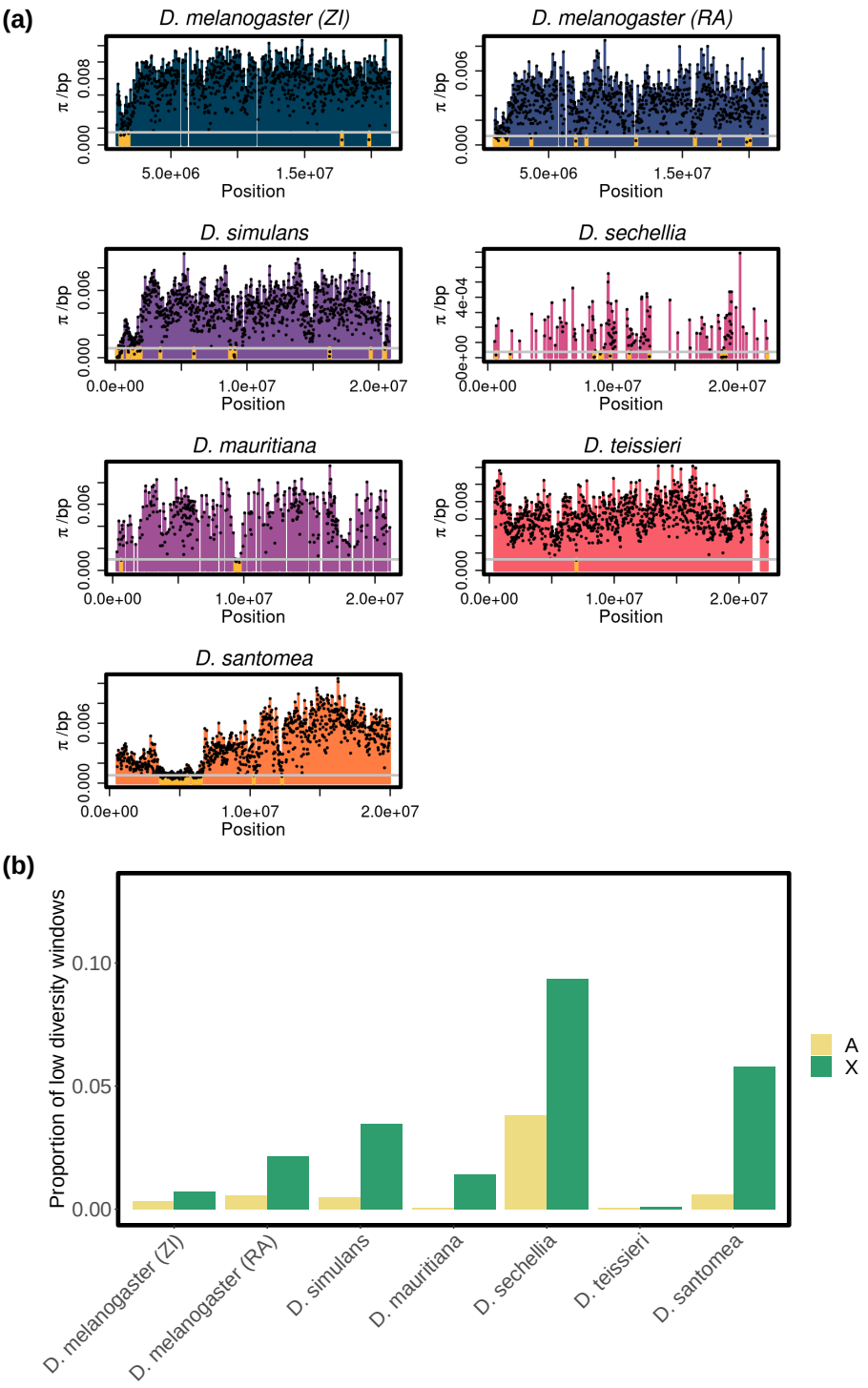

Fig. 2. Multiple *Drosophila* species have a higher proportion of windows on the X with steep depressions of diversity compared to autosomes. a) π/bp in 20-kb windows on the X chromosome of 6 *Drosophila* species and 2 *D. melanogaster* populations from RA and ZI. Windows where diversity is below 20% of the chromosomal average (gray horizontal line) are highlighted in gold. In Supplementary Fig. 4, the same figure for all autosomes is plotted for each species. b) Proportion of windows below the 20% of the chromosomal average on the autosomes (yellow) and the X chromosome (green).

mutation rate is expected when there are a higher number of germline cell divisions in males than females (Drost and Lee 1995; Kirkpatrick and Hall 2004). However, in *Drosophila*, the number of cell divisions in the male and female germlines has been shown to be similar (Drost and Lee 1995), suggesting that there should be no significant difference in the mutation rate between the sexes. Nevertheless, other factors may differentially affect X-linked and autosomal mutation rates such as the selection on codon usage on the X given a higher level of codon usage bias and GC content on the X across many *Drosophila* species (Singh *et al.* 2005a, 2005b, 2008; Vicoso *et al.* 2008; Campos *et al.* 2013; Schrider *et al.* 2013; Keightley *et al.* 2014). Past studies on *D. melanogaster* have not shown a statistically significant difference between X-linked and autosomal mutation rates (Keightley *et al.*

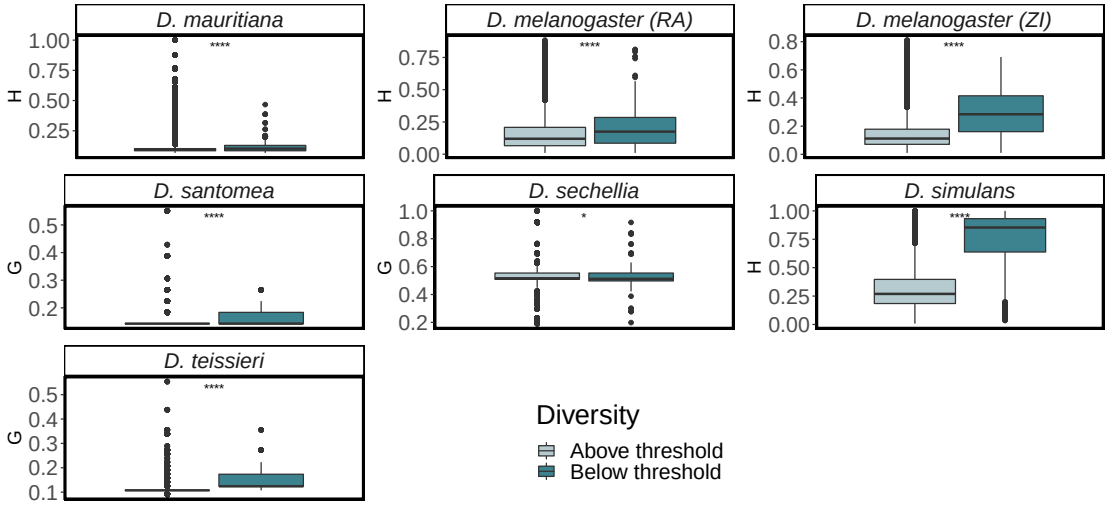

Fig. 3. Homozygosity in windows below and above 20% of the average π/bp on the X chromosome. Windows with π/bp below 20% of the chromosomal mean show elevated haplotype homozygosity (H) or elevated multilocus genotype homozygosity (G) compared to windows with π/bp above this threshold (1-sided Wilcoxon rank-sum test $P < 0.05$).

2009, 2014; Schrider *et al.* 2013). For other species, evidence of a lower X-linked mutation rate is inconclusive, with few species showing evidence of lower X mutation rates (Garrigan *et al.* 2014); however, more studies are needed to better understand X-linked vs autosomal mutation rate differences.

Nonetheless, we performed simulations to test the impact of a lower mutation rate on the X and whether this could, on its own, explain the patterns observed in the data. Simulations show that only the most extreme case of lower X-linked mutation rates ($\mu_X = 0.1\mu_A$) can decrease diversity below the 20% of $\pi_{X,\,\text{Neutral}}$, where $\pi_{X,\,\text{neutral}}$ is the average π/bp from neutral X chromosome simulations. However, none of our low X-linked mutation rate simulations significantly elevated haplotype homozygosity. In fact, the only case that can considerably reduce π_X, $\mu_X = 0.1\mu_A$, shows the lowest haplotype homozygosity across all models (Fig. 4c). This is due to longer windows in terms of base pairs for the same SNP window size used across all scenarios. The longer windows lead to a lower probability of sampling 2 identical haplotypes and hence a lower haplotype homozygosity. From the above, we conclude that it is unlikely that the low diversity windows on the X are uniquely explained by a lower X-linked mutation rate.

Demography

Next, we tested the effects of demography on X vs autosome diversity, as population bottlenecks can have different effects on the X vs autosomes due to differences in N_e. We tested 2 variations of a bottleneck model: a severe and short bottleneck and a shallow and long bottleneck (see *Methods*). Both models were fit to π/bp and S/bp in short introns from the DGRP data set (Garud *et al.* 2015, 2021). Short introns, known to evolve almost neutrally, were defined as introns shorter than 86 bp with the first 16 bp and last 6 bp removed (Clemente and Vogl 2012; Lawrie *et al.* 2013). Our simulations revealed a π_X/π_A slightly below 0.75 as well as a significant increase in haplotype homozygosity in both bottleneck scenarios (Fig. 4a and b). To test whether the higher variance on the X, expected from its smaller N_e, could result in local strong dips in diversity, we looked at whether the distribution of π_X could achieve values below 20% of the average π_X for each model. In Fig. 4d, we see that the 20% of the average π_X is well below the tail of the distribution of π_X, suggesting that bottlenecks are unlikely to generate the strong local dips in π that we observe in the data (Fig. 2).

Finally, inbreeding is another process that could result in local depletion of diversity. However, if this were the case, we would expect to see similar reductions in diversity across the autosomes, which we do not observe. Nonetheless, removal of closely related individuals in the *D. melanogaster* populations (see *Methods*) still results in depleted nucleotide diversity on the X relative to autosomes.

The effect of BGS on nucleotide diversity and haplotype homozygosity

Having investigated neutral evolutionary scenarios, we next examined whether selection can generate the patterns observed in the data. Both BGS and selective sweeps can decrease genetic diversity at linked sites, either through the purging of neutral alleles that are linked to deleterious mutations or through hitchhiking, in which a beneficial mutation and its genetic background spread rapidly through the population (Tajima 1989; Charlesworth *et al.* 1993, 1995; Stephan 2010). However, BGS is not known to increase haplotype homozygosity (Enard *et al.* 2014; Schrider 2020).

To test whether BGS can, on its own, give rise to values of $\pi_X/\pi_A < 0.75$, decrease π/bp below the low diversity threshold, and elevate haplotype homozygosity, we simulated a population ($N_e = 1 \times 10^6$) in which a fraction (d = 0.1, 0.5, and 0.8) of mutations are deleterious. The selection coefficient (s_d) for the deleterious mutations followed a gamma-distributed distribution of fitness effects (DFE) with mean and shape parameter of −0.000133 and 0.35, respectively (Huber *et al.* 2017). Our simulations show that (1) BGS does not decrease π_X such that π_X/π_A is considerably below 0.75 (Fig. 5a), (2) diversity does not dip below the 20% of $\pi_{X,\,\text{neutral}}$ threshold (Fig. 5b), and (3) none of the BGS models considered can elevate homozygosity (Fig. 5c).

The combined effect of BGS, demography, X-linked mutation rates, and recombination rate variation on nucleotide diversity and haplotype homozygosity

Separately, each of the variables we have examined thus far (sex bias, low recombination rates, reduced μ_X, bottlenecks, and BGS) cannot produce strong dips in diversity and elevated haplotype

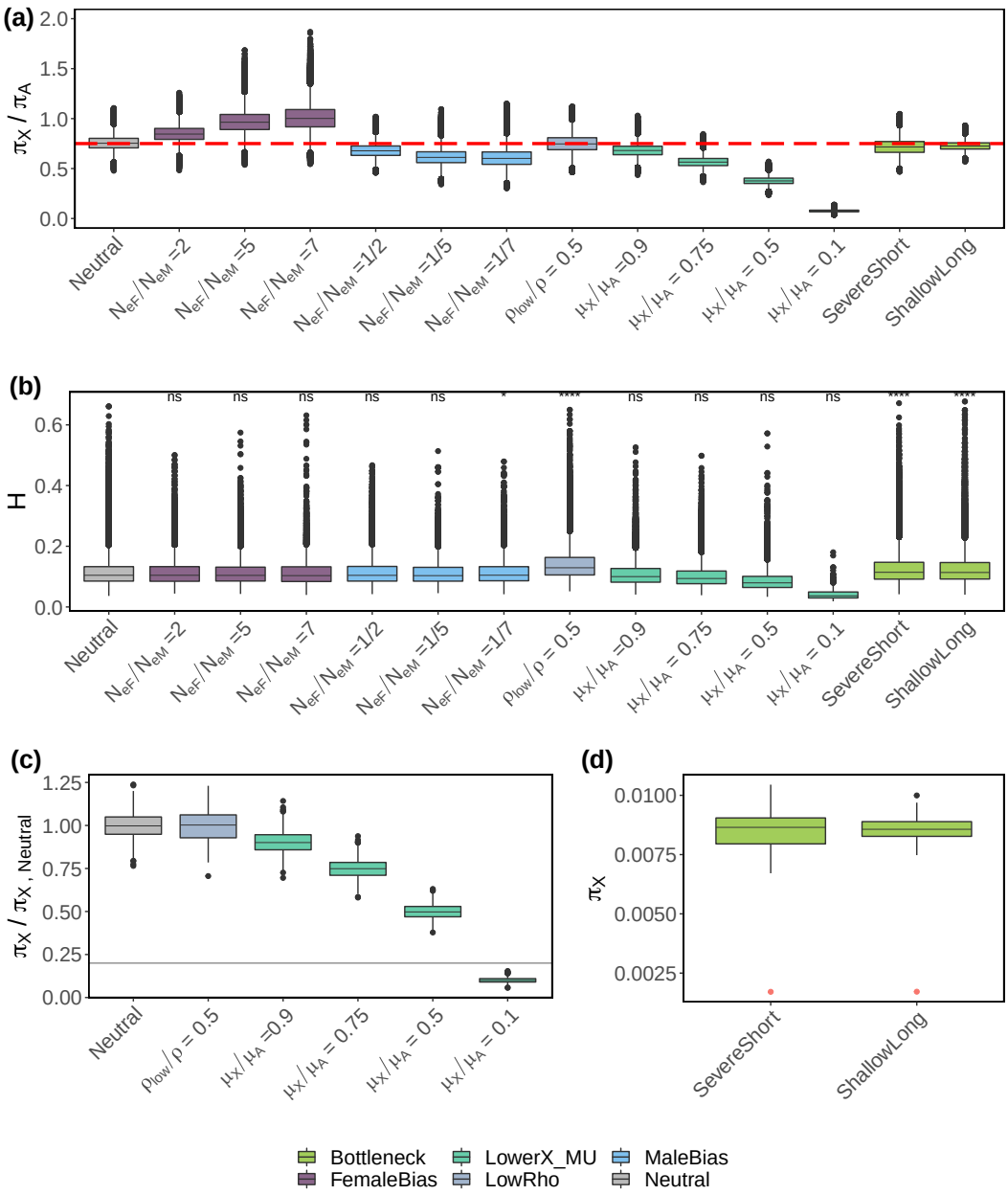

Fig. 4. Nucleotide diversity (π) and haplotype homozygosity across 12 models with no selection. The models considered include 3 female-biased models ($N_{eF}/N_{eM} = 2$, $N_{eF}/N_{eM} = 5$, and $N_{eF}/N_{eM} = 7$), 3 male-biased models ($N_{eM}/N_{eF} = 2$, $N_{eM}/N_{eF} = 5$, and $N_{eM}/N_{eF} = 7$), 1 low recombination model ($r_{low}/r = 0.5$), 4 lower X-linked mutation rate models ($\mu_X/\mu_A = 0.9$, $\mu_X/\mu_A = 0.75$, $\mu_X/\mu_A = 0.5$, and $\mu_X/\mu_A = 0.1$, where μ_X is the X-linked mutation rate and μ_A is the autosomal mutation rate), and 2 bottleneck models (a severe short bottleneck and a shallow and long bottleneck). a) π_X/π_A across all models, where the red dashed line corresponds to the expected $\pi_X/\pi_A = 0.75$ value in the baseline neutral model (same sex ratio, equal X and autosomal mutation rates, and a constant $N_e = 10^6$). b) Haplotype homozygosity across all models. The asterisks represent a significant elevation in haplotype homozygosity relative to complete neutrality using a 1-sided Wilcoxon rank-sum test. c) $\pi_X/\pi_{X,\ neutral}$ across low X mutation rate models, with $\pi_{X,\ neutral}$ equal to the average π in the baseline neutral model. The horizontal gray line is the low diversity threshold, defined as 20% of the average π/bp of the neutral case. d) π_X in 2 bottleneck models. The red data point below the boxplots corresponds to 20% of the average π from each bottleneck model.

homozygosity. However, in combination, these variables may be able to generate the patterns observed in the data. To test this, we simulated the full combination of variables, including BGS, bottlenecks, reduced μ_X, and low recombination rates, which individually showed some, but not all, of the signatures observed in the data. We did not simulate sex bias because female bias elevates diversity on the X (opposite trend of what is observed in 5 of 7 populations), and male bias is not reported to be a dominant process influencing the populations under study.

Figure 6 shows the result of combining BGS, bottlenecks, and low μ_X for $r = 5e^{-7}$ cM/bp, while Supplementary Fig. 8 shows these results for a lower recombination rate ($r = 2.5e^{-7}$ cM/bp). We found that when $r = 5e^{-7}$ cM/bp and 80% of new mutations are deleterious, less than half of the simulations displayed dips in diversity below the designated low diversity threshold. Moreover, none of these scenarios showed elevated haplotype homozygosity. With a lower recombination rate of $r = 2.5e^{-7}$ cM/bp, although haplotype homozygosity was elevated in some scenarios

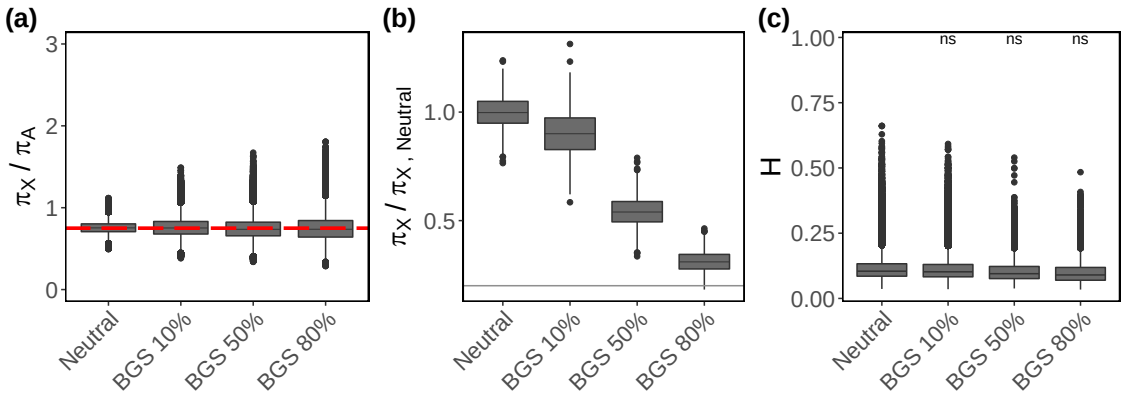

Fig. 5. Effect of BGS on diversity and haplotype homozygosity. The models considered include a neutral scenario with no sex ratio or mutation rate biases and 3 BGS models. For the BGS models, we varied the proportion of deleterious mutations (10, 50, and 80%). The DFE for the deleterious selection coefficient (s_d) was gamma distributed with mean and shape parameter of −0.000133 and 0.35, respectively (Huber *et al.* 2017). For each model, we computed a) π_X/π_A, where the red dashed line corresponds to the expected $\pi_X/\pi_A = 0.75$ value in a completely neutral case; b) $\pi_X/\pi_{X,\ neutral}$ where $\pi_{X,\ neutral}$ is the π in the constant N_e neutral model with no biases and the solid gray line is 20% of $\pi_{X,\ neutral}$; and c) haplotype homozygosity. The asterisks represent a significant elevation in haplotype homozygosity relative to neutrality using a 1-sided Wilcoxon rank-sum test.

(Supplementary Fig. 8), diversity was not sufficiently reduced to be consistent with the data.

Finally, to test whether the effect of BGS is stronger when a larger number of deleterious mutations reside on the same chromosome, we simulated longer chromosomes of 1 Mb instead of 20 kb (Supplementary Fig. 9). As before, when $r = 5e^{-7}$ cM/bp, we do not observe any of the patterns from the data. Only with a 5-fold reduction in the recombination rate ($r \leq 1e^{-8}$ cM/bp), we observe that diversity dips below 20% of $\pi_{X,\ neutral}$ and haplotype homozygosity is elevated. In this scenario, recombination is not effectively breaking linkage, leading to an elevation of haplotype homozygosity and strong reductions in diversity compared to $\pi_{X,\ neutral}$. However, in this scenario, π_X/π_A does not dip below 0.75, which is inconsistent with 5 of 7 populations in the data (Supplementary Fig. 9). Additionally, only 15% of low diversity windows in the *D. melanogaster* genome have a recombination rate below $r = 1e^{-8}$ cM/bp, making low recombination rates unlikely to be the primary force-generating dips in diversity.

Hard sweeps can generate the patterns observed on the X across species

Our results indicate that the patterns observed in the data are unlikely to be generated by sex bias, low recombination rates, low X-linked mutation rates, demography, or BGS, either individually or in combination. Next, we tested the effect of positive selection in generating dips in diversity and elevated haplotype homozygosity on the X chromosome. To do so, we simulated hard and soft sweeps (see *Methods*), varied the strength of selection ($N_e s_b$ = 20, 200, and 2,000), and computed π_X/π_A, $\pi_X/\pi_{X,\ neutral}$, and haplotype homozygosity in a 20-kb window for each scenario.

Our simulations show that selective sweeps can decrease π_X/π_A below 0.75 and elevate haplotype homozygosity (Fig. 7a and c), but only hard sweeps can reduce diversity below 20% of $\pi_{X,\ neutral}$ (Fig. 7b) as long as selection is sufficiently strong ($N_e s_b = 2{,}000$). Moreover, compared to the scenarios that we have simulated thus far, we observe a much stronger elevation in haplotype homozygosity when there is positive selection. This is indicative of a more substantial effect size, which aligns better with observations in *D. melanogaster* and *D. simulans*, the species with sample sizes comparable to those from simulations.

We note, however, that it is also possible for a strong soft sweep that is not too soft ($N_e s_b = 2{,}000$, $\theta_a = 0.1$ corresponding to approximately two sweeping haplotypes; Supplementary Fig. 10a) to reach levels of diversity below the low diversity threshold. However, in this regime ($\theta_a = 0.1$), only slightly more than half (~62%) of the sweeps simulated with $\theta_a = 0.1$ are soft (Supplementary Fig. 10), and, among those that are soft, 48% of the simulations show π_X below the low diversity threshold compared to 82% for the hard sweep model (Supplementary Fig. 11), making hard sweeps a more likely explanation of the data. Thus, the patterns of reduced diversity and elevated haplotype homozygosity on the X chromosome of the species analyzed are less likely to be solely the result of soft sweeps; rather, they align more consistently with the characteristics of hard sweeps.

Discussion

A classic question in evolutionary biology is how the evolution of the X chromosome differs from that of autosomes given the X chromosome's central role in speciation, brain function, fertility, and sexual dimorphism (Rice 1984; Saifi and Chandra 1999; Skuse 2005; Dean and Mank 2014; Payseur *et al.* 2018). Past work has suggested that the X chromosome may exhibit different evolutionary dynamics from the autosomes due to its unique inheritance pattern and increased exposure to selection through male hemizygosity (Vicoso and Charlesworth 2006; Nam *et al.* 2015; Charlesworth *et al.* 2018; Muralidhar and Veller 2022; Harris and Garud 2023). Recently, we found evidence in a North American *D. melanogaster* population that the X chromosome experiences an enrichment of hard selective sweeps compared to autosomes due to the increased visibility of new deleterious mutations to natural selection on the hemizygous X of males and reduced effective population size on the X (Harris and Garud 2023). However, it is unclear whether this pattern of enrichment of hard sweeps on the X is a universal feature across all heterogametic species.

To understand if the enrichment of hard sweeps on the X is common across species, we analyzed multiple whole-genome sequences from 6 *Drosophila* species and found evidence that suggests that hard sweeps are in fact more common on the X chromosome than autosomes across these species. Additionally, we found that the observed patterns of diversity are inconsistent

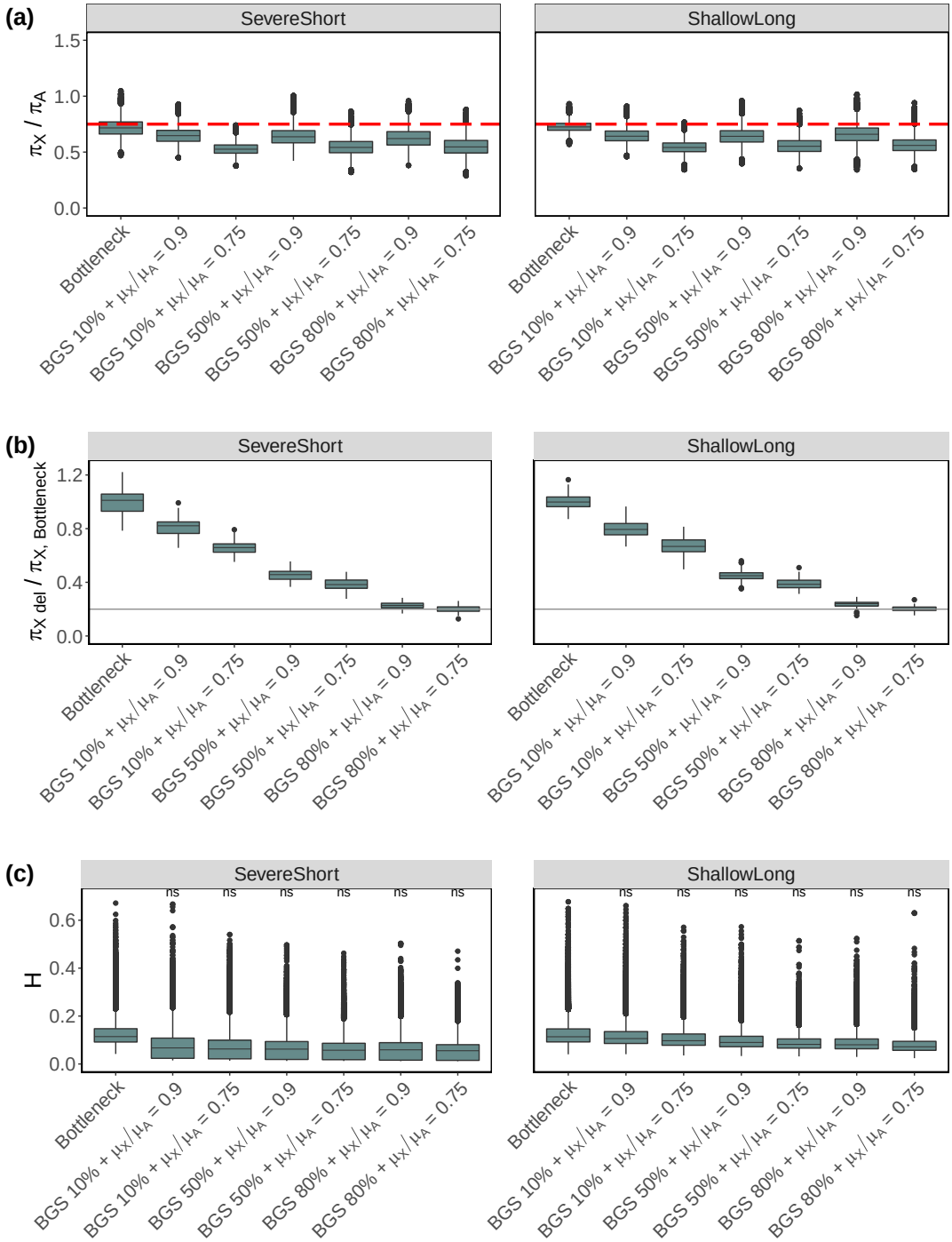

Fig. 6. Effect of BGS combined with bottlenecks and low X-linked mutation rate on diversity patterns. The models considered include the combination of a bottleneck model with 1 of 3 BGS models each combined with a lower X-linked mutation rate ($\mu_X/\mu_A = 0.9$, $\mu_X/\mu_A = 0.75$). For the BGS models, we varied the proportion of deleterious mutations (10, 50, and 80%) while the DFE for the deleterious selection coefficient (s_d) was gamma distributed with mean and shape parameter of −0.000133 and 0.35, respectively (Huber *et al.* 2017). The recombination rate was set to $r = 5e^{-7}$ cM/bp across all simulations. Supplementary Fig. 8 shows the same corresponding plot for half the recombination rate ($2.5e^{-7}$ cM/bp). The left column corresponds to a severe short bottleneck model and the right column to a shallow long bottleneck model (see *Methods*). For each model, we computed a) π_X/π_A across all models, where the red dashed line corresponds to the expected $\pi_X/\pi_A = 0.75$ value in the neutral model; b) $\pi_X/\pi_{X,\ bottleneck}$ where $\pi_{X,\ bottleneck}$ is the average π in the corresponding bottleneck model with no added biases and the solid gray line is 20% of $\pi_{X,\ bottleneck}$; and c) haplotype homozygosity across all models and results from a 1-sided Wilcoxon rank-sum test for elevation of H in the bottleneck + BGS + low μ_X scenario compared to a neutral bottleneck model.

with BGS, differences in mutation rate, population bottlenecks, or sex bias, both independently and when combined. Furthermore, we found that soft sweeps generally cannot generate the patterns observed in the data, except for when soft sweeps are not too soft (e.g. there are only 2 sweeping haplotypes). However, in this scenario ($\theta_a = 0.1$), only 62% of the simulations led to soft sweeps, and of those, only 48% generated dips in diversity falling below the low diversity threshold, whereas 82% of hard sweeps reached below this threshold. Therefore, we acknowledge that it is possible that other factors other than hard sweeps could be responsible

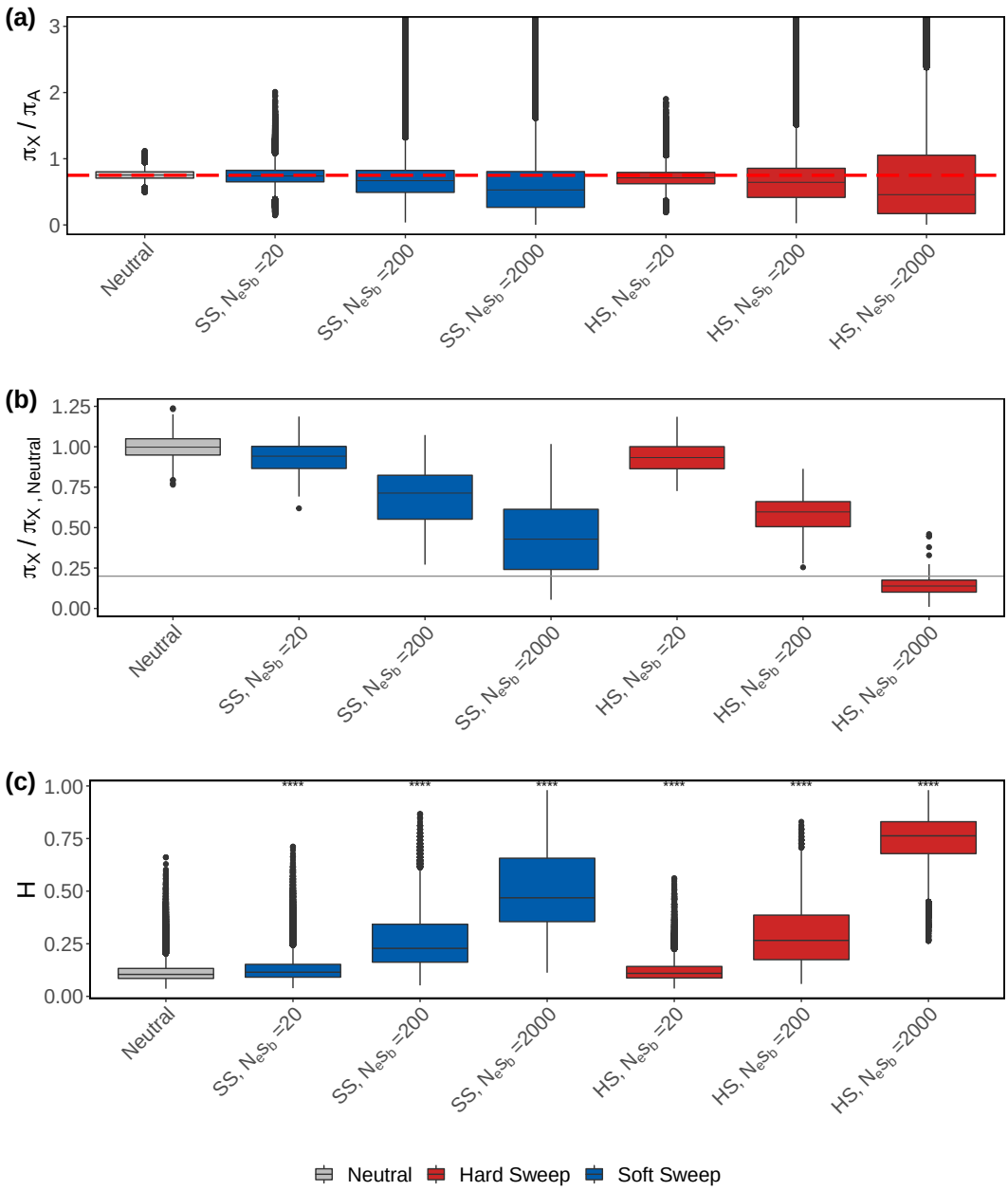

Fig. 7. Effect of hard and soft selective sweeps on diversity and haplotype homozygosity. The models considered include a neutral scenario with no sex ratio or mutation rate biases, 3 soft sweep (SS) and 3 hard sweep models (HS). We simulated soft sweep (blue) and hard sweep (red) models (see *Methods*) varying the selection strength of the adaptive mutation ($N_e s_b$ = 20, 200, and 2,000). For each model, we computed a) π_X/π_A, where the red dashed line corresponds to the expected $\pi_X/\pi_A = 0.75$ value in a completely neutral case; b) $\pi_X/\pi_{X,\,neutral}$ where $\pi_{X,\,neutral}$ is the π in the baseline neutral model and the solid gray line is 20% of $\pi_{X,\,neutral}$; and (c) haplotype homozygosity. The asterisks represent a significant elevation in haplotype homozygosity relative to complete neutrality using a 1-sided Wilcoxon rank-sum test.

for generating some of the low diversity windows observed in the data. However, our results suggest that these forces are unlikely to be the dominant processes driving the patterns in the data. Future work will be important for disentangling the effects of these forces on diversity on the X.

The finding that hard sweeps are a likely explanation for the patterns observed on the X chromosome aligns with recent theoretical work predicting harder sweeps on the X chromosome (Muralidhar and Veller 2022; Harris and Garud 2023), as well as empirical work in apes (Nam *et al.* 2015) and *D. mauritiana* (Garrigan *et al.* 2014). In our analysis, we employed a wide range of statistics, including single-site (e.g. π/bp) and multisite (e.g. haplotype homozygosity) statistics that are sensitive to signatures of selection and also leverage the additional resolution that whole-genome sequences provide over genotype data. Combining single-site and multilocus statistics is still a relatively novel area of work that has the potential to reveal patterns of evolution that cannot be detected by either type of statistic alone (Lin *et al.* 2011; Schrider and Kern 2016; Sheehan and Song 2016; Ragsdale and Gutenkunst 2017; Johri *et al.* 2020; Garud *et al.* 2021). Our approach allows us to disentangle the effects of positive selection from other forces, such as BGS and demographic processes, and rule out inconsistent models.

Comparative population genetics, in which multiple genomes from several species each are considered simultaneously, is also

a relatively new area of work given the paucity of deep population genetic samples from multiple species. Population genetic studies have traditionally focused on analyzing multiple genomes from a single species (Arbiza *et al.* 2014; Garud and Rosenberg 2015; Signor *et al.* 2018) with few examining more than 2 population genetic whole-genome data sets from different species simultaneously (Nam *et al.* 2015; Chen *et al.* 2018; Nadachowska-Brzyska *et al.* 2019; Latrille *et al.* 2023; Rodrigues *et al.* 2023 with Latrille *et al.* 2023 restricted to exomes), leaving open numerous avenues of inquiry on the commonalities and idiosyncrasies of population genetic processes.

Now, with the increasing availability of deep population genetic sequences from multiple species, we can examine population genetic processes across many species. In this study, the ability to compare multiple species reveals that the extent of positive selection on the X chromosome may not be equal across all species. For example, we observed few low diversity windows on the X in *D. teissieri*, while in *D. santomea*, we observed a low diversity region extending up to ~200 kb (Fig. 2). Additionally, this comparative approach reveals deviations from the trend in individual species. One species that in particular looked very different was *D. sechellia*. This species showed a unique trend in which π_X/π_A was the highest in exons and lowest in intergenic regions, even after stringent filtering of the data (see *Methods*; Fig. 1). A potential explanation for this behavior could be abundant introgression from *D. simulans* to *D. sechellia*, as previously reported (Garrigan *et al.* 2012; Matute and Ayroles 2014; Schrider *et al.* 2018). Interestingly, introgression appears to be less common on the X chromosome, potentially due to the involvement of sex chromosomes in hybrid incompatibilities (Maroja *et al.* 2015; Turissini and Matute 2017; Schrider *et al.* 2018). Moreover, functional regions of the genome are less likely to exhibit evidence of introgression, with exons being the least susceptible and intergenic regions having higher rates of introgressed regions (Sankararaman *et al.* 2014). As a result, if introgression is abundant, the difference in diversity between the X chromosome and autosomes should be more pronounced in intergenic regions and least in exons, potentially driving the observed trend in the data.

Despite individual species showing deviations from the trend, some species exhibited commonalities. For example, the species analyzed in this study all had a higher proportion of low diversity regions on the X chromosome compared to the autosomes. Additionally, we found some targets of selection to be shared across species; however, they were generally few in number (Supplementary Figs. 12–14; Supplementary Text 1). Instead, most genes found in low diversity windows were found in a single species only, indicating that while hard sweeps might be common on the X, different functions may be under selection in different species. We note that this finding is distinct from that of primates (Nam *et al.* 2015) where there appears to be more overlap across species. This suggests that the underlying mechanisms of selection in *Drosophila* vs primates may differ.

Our paper is a clear demonstration of how comparative population genetic analyses can reveal new insights into the forces that shape genetic variation within and across species, which we expect to serve as a useful example as new high-resolution population genomic data sets from multiple species become increasingly available in the coming years. While our study did not examine the prevalence of soft sweeps in autosomes and the X chromosome due to limited sample sizes per species, future research with larger sample sizes could provide the relevant data needed to be able to detect soft sweeps with haplotype homozygosity statistics, thereby providing a more comprehensive understanding of the tempo and mode of adaptation across species (Pennings and Hermisson 2006b). Ultimately, our study highlights the significance of hard sweeps in shaping the diversity patterns of the X chromosome across species and suggests an important evolutionary mechanism that may be widespread among all species. Future work may reveal the potential role of these hard sweeps in driving sexual dimorphism and speciation given the X chromosome's significant involvement in these important processes.

Data availability

Sequence data for *D. melanogaster* are available at www.johnpool.net, where the DGRP and DPGP3 seq files are available for download. For *D. simulans*, the vcf file can be downloaded from https://zenodo.org/record/154261#.YzMzty2z3jC. For the remaining species, sequence data are available at NCBI's Short Read Archive with accession numbers given in Supplementary Table 1. Code used for simulations as well as to process and analyze the data is available online (https://github.com/garudlab/DrosCrossSpecies_XchrHardSweeps).

Supplemental material available at GENETICS online.

Acknowledgments

We are grateful to Rebekah Rogers and Peter Andolfatto for providing the *D. simulans* w501 genome assembly. We thank Dmitri Petrov and Kirk Lohmueller for helpful conversations.

Funding

MH was supported by the National Institutes of Health grants NIH-NIGMS 5T32GM008185 and NIH T32HG002536. BYK was supported by NIH grant NIGMS F32GM135998. NRG was supported by funding from a National Science Foundation CAREER award.

Conflicts of interest

The authors declare no conflict of interest.

Literature cited

Arbiza L, Gottipati S, Siepel A, Keinan A. 2014. Contrasting X-linked and autosomal diversity across 14 human populations. Am J Hum Genet. 94(6):827–844. doi:10.1016/j.ajhg.2014.04.011.

Assaf ZJ, Petrov DA, Blundell JR. 2015. Obstruction of adaptation in diploids by recessive, strongly deleterious alleles. Proc Natl Acad Sci U S A. 112(20):E2658–E2666. doi:10.1073/pnas.1424949112.

Betancourt AJ, Kim Y, Orr HA. 2004. A pseudohitchhiking model of X vs. autosomal diversity. Genetics. 168(4):2261–2269. doi:10.1534/genetics.104.030999.

Campos JL, Zeng K, Parker DJ, Charlesworth B, Haddrill PR. 2013. Codon usage bias and effective population sizes on the X chromosome versus the autosomes in *Drosophila melanogaster*. Mol Biol Evol. 30(4):811–823. doi:10.1093/molbev/mss222.

Charlesworth B, Campos JL, Jackson BC. 2018. Faster-X evolution: theory and evidence from Drosophila. Mol Ecol. 27(19):3753–3771. doi:10.1111/mec.14534.

Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. Genetics. 141(4):1619–1632. doi:10.1093/genetics/141.4.1619.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 130(1): 113–146. doi:10.1086/284701.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. 134(4):1289–1303. doi:10.1093/genetics/134.4.1289.

Chen ZH, Zhang M, Lv FH, Ren X, Li WR, Liu MJ, Nam K, Bruford MW, Li MH. 2018. Contrasting patterns of genomic diversity reveal accelerated genetic drift but reduced directional selection on X-chromosome in wild and domestic sheep species. Genome Biol Evol. 10(5):1282–1297. doi:10.1093/gbe/evy085.

Clemente F, Vogl C. 2012. Unconstrained evolution in short introns? —an analysis of genome-wide polymorphism and divergence data from *Drosophila*. J Evol Biol. 25(10):1975–1990. doi:10.1111/j.1420-9101.2012.02580.x.

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 8(10): e1002905. doi:10.1371/journal.pgen.1002905.

Coyne JA, Orr HA. 1998. The evolutionary genetics of speciation. Philos Trans R Soc Lond B Biol Sci. 353(1366):287–305. doi:10.1098/rstb.1998.0210.

Dean R, Mank JE. 2014. The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. J Evol Biol. 27(7):1443–1453. doi:10.1111/jeb.12345.

Dieringer D, Nolte V, Schlötterer C. 2005. Population structure in African Drosophila melanogaster revealed by microsatellite analysis. Mol Ecol. 14(2):563–573. doi:10.1111/j.1365-294X.2004.02422.x.

Drost JB, Lee WR. 1995. Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. Environ Mol Mutagen. 25(Suppl. 2):48–64. doi:10.1002/em.2850250609.

Enard D, Messer PW, Petrov DA. 2014. Genome-wide signals of positive selection in human evolution. Genome Res. 24(6):885–895. doi:10.1101/gr.164822.113.

Feder AF, Pennings PS, Petrov DA. 2021. The clarifying role of time series data in the population genetics of HIV. PLoS Genet. 17(1): e1009050. doi:10.1371/journal.pgen.1009050.

Fraïsse C, Sachdeva H. 2021. The rates of introgression and barriers to genetic exchange between hybridizing species: sex chromosomes vs autosomes. Genetics. 217(2):iyaa025. doi:10.1093/GENETICS/IYAA025.

Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton KR, Presgraves DC. 2012. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. Genome Res. 22(8): 1499–1511. doi:10.1101/gr.130922.111.

Garrigan D, Kingan SB, Geneva AJ, Vedanayagam JP, Presgraves DC. 2014. Genome diversity and divergence in *Drosophila mauritiana*: multiple signatures of faster X evolution. Genome Biol Evol. 6(9):2444–2458. doi:10.1093/gbe/evu198.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. PLoS Genet. 11(2):e1005004. doi:10.1371/journal.pgen.1005004.

Garud NR, Messer PW, Petrov DA. 2021. Detection of hard and soft selective sweeps from *Drosophila melanogaster* population genomic data. PLoS Genet. 17(2):e1009373. doi:10.1371/journal.pgen.1009373.

Garud NR, Rosenberg NA. 2015. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. Theor Popul Biol. 102:94–101. doi:10.1016/j.tpb.2015.04.001.

Gillespie JH. 2004. Population Genetics: A Concise Guide. Baltimore (MD): Johns Hopkins University Press.

Glaser-Schmitt A, Wittmann MJ, Ramnarine TJS, Parsch J. 2021. Sexual antagonism, temporally fluctuating selection, and variable dominance affect a regulatory polymorphism in *Drosophila melanogaster*. Mol Biol Evol. 38(11):4891–4907. doi:10.1093/molbev/msab215.

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, Dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, *et al.* 2022. FlyBase: a guided tour of highlighted features. Genetics. 220(4): iyac035. doi:10.1093/genetics/iyac035.

Haller BC, Messer PW. 2019. SLim 3: forward genetic simulations beyond the Wright–Fisher model. Mol Biol Evol. 36(3):632–637. doi:10.1093/molbev/msy228.

Harris M, Garud NR. 2023. Enrichment of hard sweeps on the X chromosome in *Drosophila melanogaster*. Mol Biol Evol. 40(1): msac268. doi:10.1093/molbev/msac268.

Harris AM, Garud NR, DeGiorgio M. 2018. Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. Genetics. 210(4):1429–1452. doi:10.1534/genetics.118.301502.

Harris RB, Sackman A, Jensen JD. 2018. On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. PLoS Genet. 14(12):e1007859. doi:10.1371/journal.pgen.1007859.

Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 169(4):2335–2352. doi:10.1534/genetics.104.036947.

Hermisson J, Pennings PS. 2017. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. Methods Ecol Evol. 8(6):700–716. doi:10.1111/2041-210X.12808.

Huber CD, Kim BY, Marsden CD, Lohmueller KE. 2017. Determining the factors driving selective effects of new nonsynonymous mutations. Proc Natl Acad Sci U S A. 114(17):4465–4470. doi:10.1073/pnas.1619508114.

Johri P, Charlesworth B, Jensen JD. 2020. Toward an evolutionarily appropriate null model: jointly inferring demography and purifying selection. Genetics. 215(1):173–192. doi:10.1534/genetics.119.303002.

Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. Genetics. 123(4):887–899. doi:10.1093/genetics/123.4.887.

Kauer M, Zangerl B, Dieringer D, Schlötterer C. 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. Genetics. 160(1):247–256. doi:10.1093/genetics/160.1.247.

Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics. 196(1):313–320. doi:10.1534/genetics.113.158758.

Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 19(7):1195–1201. doi:10.1101/gr.091231.109.

Kirkpatrick M, Hall DW. 2004. Male-biased mutation, sex linkage, and the rate of adaptive evolution. Evolution. 58(2):437–440. https://doi.org/10.1111/j.0014-3820.2004.tb01659.x.

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. 2023. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 51(D1):D445–D451. doi:10.1093/nar/gkac998.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015. The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range

population. Genetics. 199(4):1229–1241. doi:10.1534/genetics.115.174664.

Latrille T, Rodrigue N, Lartillot N. 2023. Genes and sites under adaptation at the phylogenetic scale also exhibit adaptation at the population-genetic scale. Proc Natl Acad Sci U S A. 120(11): e2214977120. doi:10.1073/pnas.2214977120.

Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at synonymous sites in *D. melanogaster*. PLoS Genet. 9(5):e1003527. doi:10.1371/journal.pgen.1003527.

Legrand D, Tenaillon MI, Matyot P, Gerlach J, Lachaise D, Cariou ML. 2009. Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. Genetics. 182(4):1197–1206. doi:10.1534/genetics.108.092080.

Levine RP. 1955. Chromosome structure and the mechanism of crossing over. Proc Natl Acad Sci U S A. 41(10):727–730. doi:10.1073/pnas.41.10.727.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv 1303.3997. https://doi.org/10.48550/arXiv.1303.3997, preprint: not peer reviewed.

Lin K, Li H, Schlötterer C, Futschik A. 2011. Distinguishing positive selection from neutral evolution: boosting the performance of summary statistics. Genetics. 187(1):229–244. doi:10.1534/genetics.110.122614.

Maroja LS, Larson EL, Bogdanowicz SM, Harrison RG. 2015. Genes with restricted introgression in a field cricket (*Gryllus firmus/Gryllus pennsylvanicus*) hybrid zone are concentrated on the X chromosome and a single autosome. G3 (Bethesda). 5(11): 2219–2227. doi:10.1534/g3.115.021246.

Mather K. 1939. Crossing over and heterochromatin in the X chromosome of *Drosophila melanogaster*. Genetics. 24(3):413–435. doi:10.1093/genetics/24.3.413.

Matute DR, Ayroles JF. 2014. Hybridization occurs between *Drosophila simulans* and *D. sechellia* in the Seychelles archipelago. J Evol Biol. 27(6):1057–1068. doi:10.1111/jeb.12391.

McGrath C. 2022. Highlight: comparative population genomics—answering old questions with new data. Genome Biol Evol. 14(1): evab278. doi:10.1093/gbe/evab278.

McVicker G, Gordon D, Davis C, Green P. 2009. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5(5):e1000471. doi:10.1371/journal.pgen.1000471.

Meany MK, Conner WR, Richter SV, Bailey JA, Turelli M, Cooper BS. 2019. Loss of cytoplasmic incompatibility and minimal fecundity effects explain relatively low Wolbachia frequencies in *Drosophila mauritiana*. Evolution. 73:1278–1295. doi:10.1111/evo.13745.

Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol. 28(11):659–669. doi:10.1016/j.tree.2013.08.003.

Muirhead CA, Presgraves DC. 2016. Hybrid incompatibilities, local adaptation, and the genomic distribution of natural introgression between species. Am Nat. 187(2):249–261. doi:10.1086/684583.

Muralidhar P, Veller C. 2022. Dominance shifts increase the likelihood of soft selective sweeps. Evolution. 76(5):966–984. doi:10.1111/evo.14459.

Nadachowska-Brzyska K, Burri R, Ellegren H. 2019. Footprints of adaptive evolution revealed by whole Z chromosomes haplotypes in flycatchers. Mol Ecol. 28(9):2290–2304. doi:10.1111/mec.15021.

Nam K, Munch K, Hobolth A, Dutheil JY, Veeramah KR, Woerner AE, Hammer MF; Great Ape Genome Diversity Project; Mailund T, Schierup MH. 2015. Extreme selective sweeps independently targeted the X chromosomes of the great apes. Proc Natl Acad Sci U S A. 112(20):6413–6418. doi:10.1073/pnas.1419306112.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, *et al.* 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44(D1):D733–D745. doi:10.1093/nar/gkv1189.

Orr HA, Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics. 157(2):875–884. doi:10.1093/genetics/157.2.875.

Patten MM. 2019. The X chromosome favors males under sexually antagonistic selection. Evolution. 73:84–91. doi:10.1111/evo.13646.

Payseur BA, Presgraves DC, Filatov DA. 2018. Introduction: Sex chromosomes and speciation. Molecular Ecology. doi: 10.1111/mec.14828.

Pennings PS, Hermisson J. 2006a. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol. 23(5):1076–1084. doi:10.1093/molbev/msj117.

Pennings PS, Hermisson J. 2006b. Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet. 2(12): 1998–2012. doi:10.1371/journal.pgen.0020186.

Peter BM, Huerta-Sanchez E, Nielsen R. 2012. Distinguishing between selective sweeps from standing variation and from a de novo mutation. PLoS Genet. 8(10):e1003011. doi:10.1371/journal.pgen.1003011.

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P, Emerson JJ, Saelao P, Begun DJ, *et al.* 2012. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet. 8(12):e1003080. doi:10.1371/journal.pgen.1003080.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842. doi:10.1093/bioinformatics/btq033.

Ragsdale AP, Gutenkunst RN. 2017. Inferring demographic history using two-locus statistics. Genetics. 206(2):1037–1048. doi:10.1534/genetics.117.201251.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution. 38(4):735–742. doi:10.1111/j.1558-5646.1984.tb00346.x.

Rodrigues MF, Kern AD, Ralph PL. 2023. Shared evolutionary processes shape landscapes of genomic variation in the great apes. bioRxiv 527547. https://doi.org/10.1101/2023.02.07.527547, preprint: not peer reviewed.

Rogers RL, Shao L, Sanjak JS, Andolfatto P, Thornton KR. 2014. Revised annotations, sex-biased expression, and lineage-specific genes in the *Drosophila melanogaster* group. G3 (Bethesda). 4(12): 2345–2351. doi:10.1534/g3.114.013532.

Saifi GM, Chandra HS. 1999. An apparent excess of sex and reproduction-related genes on the human X chromosome. Proc R Soc B Biol Sci. 266(1415):203–209. doi:10.1098/rspb.1999.0623.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 507(7492):354–357. doi:10.1038/nature12961.

Schrider DR. 2020. Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. Genetics. 216(2):499–519. doi:10.1534/genetics.120.303469.

Schrider DR, Ayroles J, Matute DR, Kern AD. 2018. Supervised machine learning reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. PLoS Genet. 14(4):e1007341. doi:10.1371/JOURNAL.PGEN.1007341.

Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. Genetics. 194(4):937–954. doi:10.1534/genetics.113.151670.

Schrider DR, Kern AD. 2016. S/HIC: robust identification of soft and hard sweeps using machine learning. PLoS Genet. 12(3): e1005928. doi:10.1371/journal.pgen.1005928.

Schrider DR, Mendes FK, Hahn MW, Kern AD. 2015. Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. Genetics. 200(1):267–284. doi:10.1534/genetics.115.174912.

Serrato-Capuchina A, D'Agostino ERR, Peede D, Roy B, Isbell K, Wang J, Matute DR. 2021. P-elements strengthen reproductive isolation within the *Drosophila simulans* species complex. Evolution. 75(10): 2425–2440. doi:10.1111/evo.14319.

Sheehan S, Song YS. 2016. Deep learning for population genetic inference. PLoS Comput Biol. 12(3):e1004845. doi:10.1371/journal.pcbi.1004845.

Signor SA, New FN, Nuzhdin S. 2018. A large panel of *Drosophila simulans* reveals an abundance of common variants. Genome Biol Evol. 10(1):189–206. doi:10.1093/gbe/evx262.

Singh ND, Davis JC, Petrov DA. 2005a. X-linked genes evolve higher codon bias in *Drosophila* and *Caenorhabditis*. Genetics. 171(1): 145–155. doi:10.1534/genetics.105.043497.

Singh ND, Davis JC, Petrov DA. 2005b. Codon bias and noncoding GC content correlate negatively with recombination rate on the *Drosophila* X chromosome. J Mol Evol. 61(3):315–324. doi:10.1007/s00239-004-0287-1.

Singh ND, Larracuente AM, Clark AG. 2008. Contrasting the efficacy of selection on the X and autosomes in *Drosophila*. Mol Biol Evol. 25(2):454–467. doi:10.1093/molbev/msm275.

Singh ND, Macpherson JM, Jensen JD, Petrov DA. 2007. Similar levels of X-linked and autosomal nucleotide variation in African and non-African populations of *Drosophila melanogaster*. BMC Evol Biol. 7(1):202. doi:10.1186/1471-2148-7-202.

Skuse DH. 2005. X-linked genes and mental functioning. Hum Mol Genet. 14(Suppl. 1):R27–R32. doi:10.1093/hmg/ddi112.

Smit AFA, Hubley R, Grenn P. 2013-2015. RepeatMasker Open-4.0. Available from http://www.repeatmasker.org.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res (Camb). 23(1):23–35. doi:10.1017/S0016672300014634.

Stephan W. 2010. Genetic hitchhiking versus background selection: the controversy and its implications. Philos Trans R Soc B Biol Sci. 365(1544):1245–1253. doi:10.1098/rstb.2009.0278.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123(3):585–595. doi:10.1093/genetics/123.3.585.

Thornton K, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics. 172(3): 1607–1619. doi:10.1534/genetics.105.048223.

Turelli M, Orr HA. 1995. The dominance theory of HALDANE's rule. Genetics. 140(1):389–402. doi:10.1093/genetics/140.1.389.

Turissini DA, Matute DR. 2017. Fine scale mapping of genomic introgressions within the *Drosophila yakuba* clade. PLoS Genet. 13(9): e1006971. doi:10.1371/journal.pgen.1006971.

Van der Auwera GA, O'Connor BD. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O' Riley Media.

Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet. 7(8):645–653. doi:10.1038/nrg1914.

Vicoso B, Charlesworth B. 2009. Effective population size and the faster-X effect: an extended model. Evolution. 63(9):2413–2426. doi:10.1111/j.1558-5646.2009.00719.x.

Vicoso B, Haddrill PR, Charlesworth B. 2008. A multispecies approach for comparing sequence evolution of X-linked and autosomal sites in *Drosophila*. Genet Res. 90(5):421–431. doi:10.1017/S0016672308009804.

Vincenten N, Kuhl LM, Lam I, Oke A, Kerr AR, Hochwagen A, Fung J, Keeney S, Vader G, Marston AL. 2015. The kinetochore prevents centromere-proximal crossover recombination during meiosis. eLife. 4:e10850. doi:10.7554/eLife.10850.

Wall JD, Pritchard JK. 2003. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 4(8):587–597. doi:10.1038/nrg1123.

Editor: M. Hahn