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

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

[1] Department of Computational Medicine, University of California Los Angeles, Los Angeles California, United States of America
[2] Department of Biology, Stanford University, Stanford, California, United States of America
[3] Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, United States of America
[4] Department of Human Genetics, University of California, Los Angeles, California, United States of America

*Correspondence to ngarud@ucla.edu

Abstract

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 six 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.

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 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 two 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 two 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 (Maynard 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 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 versus soft sweeps in natural populations has been of great interest and debate (Peter *et al.* 2012; Assaf *et al.* 2015; Schrider *et al.* 2015; Schrider and Kern 2016; Harris *et al.* 2018b; 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 six *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 (Garud *et al.* 2015; Harris and Garud 2023) used in our previous work (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 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 six *Drosophila* species from seven populations (*D. melanogaster* from Zambia (ZI), *D. melanogaster* from Raleigh (RA)*, D. simulans, D. sechellia, D. mauritiana, D. santomea* and *D. teissieri*), which we downloaded and processed as follows:

92 For both *D. melanogaster* populations, data are publicly available as part of the Drosophila
93 Genome Nexus data set (Lack *et al.* 2015) and can be downloaded from www.johnpool.net. We
94 downloaded 205 DGRP genomes from Raleigh (RA), North Carolina and 197 DPGP3 genomes
95 from Zambia (ZI). For our analysis we used 100 genomes from each population, previously
96 processed (Harris and Garud 2023) to remove individuals having high IBD with one another, as
97 well as residual heterozygosity, and high levels of missing data.
98 For *D. simulans* we downloaded data from 170 inbred lines from a North American
99 population (Signor *et al.* 2018) available at https://zenodo.org/record/154261#.YzMzty2z3jC. For
100 the remaining four species, we obtained genomes from NCBI's RefSeq and short read data from
101 NCBI's Short Read Archive (Garrigan *et al.* 2014; Turissini and Matute 2017; Meany *et al.* 2019;
102 Serrato-Capuchina *et al.* 2021) (**Table S1**).
103 For *D. mauritiana , D.sechellia, D. teissieri* and *D. santomea* variant calling was performed
104 with the NVIDIA Clara Parabricks pipeline v4.0.0
105 (https://docs.nvidia.com/clara/parabricks/4.0.0/index.html), a re-implementation of variant calling
106 tools including BWA-mem (Li 2013) and GATK4 (Van der Auwera and O'Connor 2020)
107 optimized for running on NVIDIA graphics processing units. Following the GATK best practices
108 for germline variant calling (https://gatk.broadinstitute.org/hc/en-us/sections/360007226651), we
109 mapped each sample's reads to the appropriate reference genome, sorted and removed PCR
110 duplicates, generated single-sample GVCFs, then performed joint genotyping with
111 HaplotypeCaller. From this initial set of variant calls, we removed sites with quality scores
112 QUAL<30.0 to obtain a bootstrap set of high-confidence calls for another round of variant calling
113 (https://gatk.broadinstitute.org/hc/en-us/articles/360035890531-Base-Quality-Score-
114 Recalibration-BQSR-). We performed base quality score recalibration on the mapped reads using
115 the bootstrap set then once again generated single sample GVCFs that were used for joint
116 genotyping with HaplotypeCaller to ultimately generate a VCF file for each species.
117 To remove any low quality variant sites, we applied GATK recommended hard filters
118 (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-
119 variants), summarized as follows: FS > 60, QD < 2.0, MQ < 4, MQRankSum < -12.5, QUAL <
120 30.0, SOR > 3.0, and ReadPosRankSum < -8.0. Additionally, we excluded sites that were not
121 uniquely mappable (**Fig. S1**). We also filtered out sites lying within repetitive elements as

122 predicted by RepeatMasker (Smit *et al.* 2015). Only bi-allelic SNPs were considered in our
123 analysis.
124 Next, we excluded invariant sites of poor quality. To do so we filtered entire regions, which
125 included both variant and invariant sites, with poor depth and quality statistics. We used banded
126 gVCF files, where sites of similar quality get concatenated into a band. We excluded intervals that
127 failed to meet the depth and quality criteria: ('MIN(FMT/DP)>10 & MIN(FMT/GQ)>30'). Next,
128 we combined the results from the GATK hard filters applied to variant sites with the resulting
129 regions – comprising both variant and invariant sites – that passed the filter applied to the banded
130 gVFC files. This gave us the total number of callable sites in the genome. When computing
131 statistics using a sliding window approach (see section below), we excluded windows that
132 overlapped 50% more with low-quality intervals.
133 High-density repeat regions are often associated with centromeric regions, which consist
134 of highly homogeneous tandem repeats and are known to experience low rates of recombination
135 (Mather 1939; Levine 1955; Vincenten *et al.* 2015). Consequently, these regions tend to exhibit
136 reduced diversity and increased homozygosity, which can lead to false positive signals of selection.
137 To mitigate potential confounding effects of repeats on selection inferences, we calculated the
138 proportion of base pairs identified as repeats within 50Kb windows. In addition to removing
139 individual sites masked as repeats, we removed entire windows in which 20% or more of the sites
140 were marked as repeats by RepeatMasker.
141
142 **Diversity statistics and haplotype homozygosity measured from data**
143 We annotated which SNPs lie in exon, intron or intergenic regions using RefSeq (O'Leary
144 *et al.* 2016) annotations. The corresponding assembly accession numbers for each species are:
145 GCF_004382145.1 (*D. mauritiana)*, GCF_004382195.2 (*D. sechellia*), GCF_016746235.2 (*D.*
146 *teissieri*), GCF_016746245.2 (*D. santomea*) and GCF_000001215.2 (*D. melanogaster*). For *D.*
147 *simulans* we used the reference genome and corresponding gff file provided by Rebekah Rogers
148 and Peter Andolfatto (Rogers *et al.* 2014). We used BEDtools v.2.3.0 (Quinlan and Hall 2010) to
149 separate the data into exon, introns or intergenic regions. If any position could be included in
150 multiple categories, we annotated the position prioritizing exons followed by introns and finally
151 intergenic regions. Additionally, we excluded regions less than 40bp long.

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

159 To identify regions of the genome that may have exceptionally low diversity due to
160 selection, we also computed π/bp in non-overlapping windows of 10, 20 and 50 kb (**Fig. S2**) across
161 the autosomes and the X chromosome of all species. We then defined a low diversity threshold as
162 X% of the π/bp average of each chromosome with X=20 or 30%. We labeled windows with π/bp
163 below this threshold as low diversity windows, further investigated as putative sweep regions. For
164 the main analysis of this work, 20Kb windows and a threshold of 20% of the average π/bp was
165 used. To test for differences in quality in windows below and above the defined threshold, we
166 computed the proportion of missing data and quality score per site and found similar distributions
167 across these categories (**Fig. S3**).

168 Additionally, we computed haplotype or multi-locus genotype homozygosity in windows
169 of 20 SNPs long on the X chromosome, with the expectation that haplotype homozygosity should
170 be elevated in selective sweeps but not background selection (Wall and Pritchard 2003; Enard *et*
171 *al.* 2014; Garud *et al.* 2015; Schrider 2020). In species for which we had phased data (*D.*
172 *melanogaster* and *D. simulans*) we computed the expected haplotype homozygosity (Garud *et al.*
173 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
174 sample with n distinct haplotypes. For the remaining species we computed the expected multi-
175 locus genotype homozygosity (Harris *et al.* 2018a) defined as $G = \sum_{i=1}^{n} q_i^2$, where q_i is the
176 frequency of the ith most common multi-locus genotype in a sample with n distinct haplotypes. In
177 contrast to a phased haplotype, where the allelic state for each site is known, a multi-locus genotype
178 is a string that represents the diploid state of the individual where each site is labeled as
179 either homozygous for the reference allele, homozygous for the alternate allele, or heterozygous.
180 A high recombination rate is expected to break haplotypes and thus decrease both H and G,
181 whereas with a low recombination rate, haplotypes may drift to high frequency and thus increase
182 H, and by extension G.

183 In addition to measuring H and G in 20 SNP windows, we tested windows of length 50 and
184 10 SNPs long. However, due to the small sample size of most species, the probability of observing
185 the same multi-locus genotype twice becomes small with longer window sizes, making it difficult
186 to capture any signal in the 50 SNP window case. With 10 SNP windows we also found an
187 elevation of homozygosity in low diversity windows. However, we opted for 20 SNP windows as
188 smaller window sizes may increase homozygosity due to genetic drift.
189
190 **Simulation analysis**
191 To understand the evolutionary processes responsible for the patterns of π/bp and haplotype
192 homozygosity observed in the data, we simulated a variety of evolutionary models for the X
193 chromosome and autosomes using SLiM 3.7 (Haller and Messer 2019). The scenarios simulated
194 included neutrality, sex bias, low recombination rate, mutation rate bias, bottlenecks, background
195 selection, hard sweeps and soft sweeps as described below.
196 For all models with a fixed population size (excluding bottleneck models) we simulate a
197 constant $Ne=10^6$ population. An N_e of this order of magnitude is reasonable for most of the species
198 analyzed given mean nucleotide diversity levels, with exception of *D. sechellia*, which has been
199 shown to have a lower $Ne\sim10^5$ (Legrand *et al.* 2009). SLiM is a forward in time simulator, which
200 makes simulating populations with $N_e > 5x10^5$ unfeasible due to memory requirements. As in our
201 previous work (Harris and Garud 2023) we rescaled our simulations using a constant factor of
202 Q=50.
203 In all our simulations we modeled a 20kb region with a recombination rate of $r = 5x10^{-7}$
204 cM/bp (unless otherwise specified) and a neutral mutation rate of $\mu=1x10^{-9}$, both rescaled by Q=50.
205 For the simulations that include selection we assumed that mutations on the X of males experience
206 the same fitness effect as that of a homozygous female (i.e. dosage compensation). We ran a total
207 of 100 simulations for each model and ran a $10N_e$ burn-in for every simulation.
208
209 *Neutral models*
210 We first simulated a completely neutral model with a female to male ratio of one and equal
211 mutation rates between the sexes. Next, we introduced scenarios that can differentially impact X-
212 linked and autosomal diversity. We simulated female and male sex bias varying the sex ratio as
213 2:1, 5:1 and 7:1 for each scenario, as well as a model with a lower X-linked mutation rate reducing

214 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
215 tested the effect of regions of low recombination by running simulations with recombination
216 reduced to 1,10, 20 and 50% of the original *r*.

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

223

224 *Background selection*

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

236

237 *Selective sweeps*

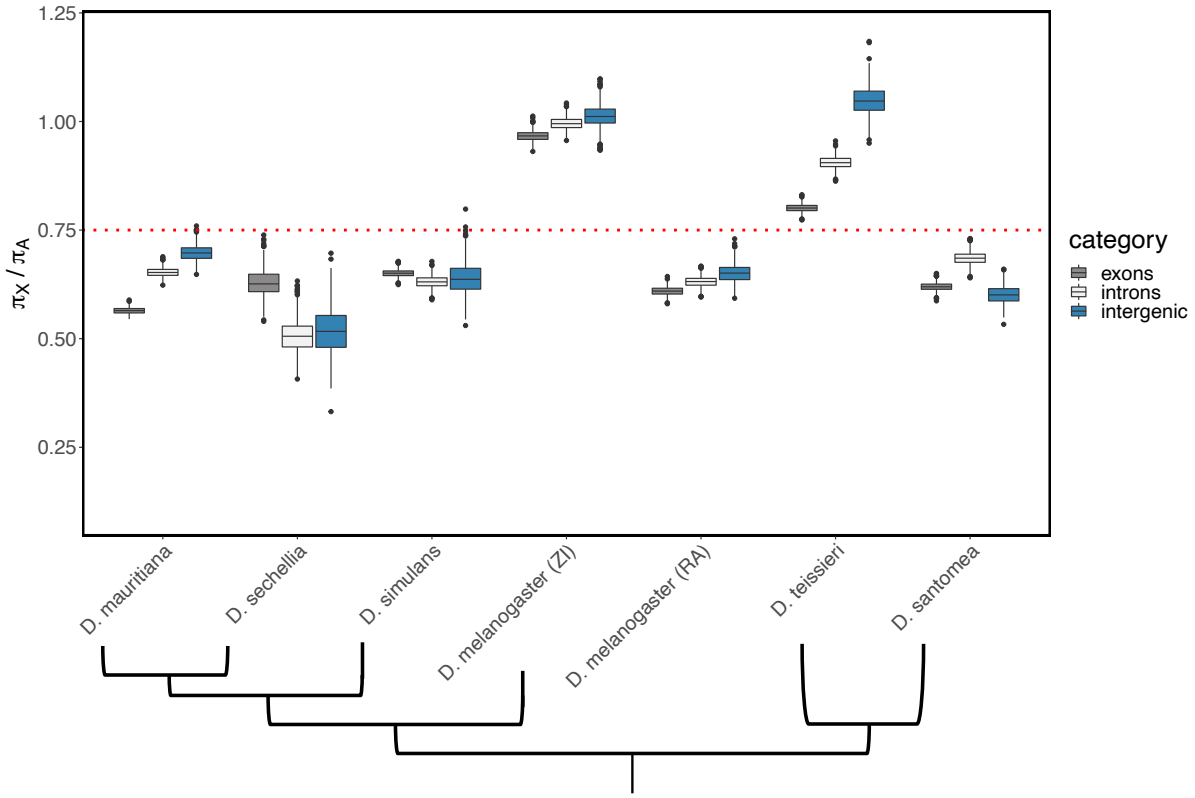
238 To understand the effects of positive selection on patterns of diversity on the X
239 chromosome and autosomes we simulated hard and soft sweeps. To model hard sweeps, we
240 introduced a single adaptive mutation on the center of the haplotype (θ_a=0.01) and restarted the
241 simulation if the adaptive mutation was lost. We conditioned on fixation of the sweep and sampled
242 100 haplotypes.

243 To simulate soft sweeps, we introduced adaptive mutations recurrently at the center of the
244 haplotype at a rate defined by θ_a=0.1, 1 and 10, where $\theta_a=4Ne\mu_a$ with μ_a the adaptive mutation

245 rate. We conditioned our simulations on the fixation of the sweep, after which we took a sample
246 of n=100 haplotypes. We verified that our simulation represented a soft sweep by only including
247 samples with two or more mutational origins in our analysis. For both hard and soft sweeps, we
248 varied the strength of selection such that Nes_b=20, 200 or 2000.
249
250 **Identification of shared genes under selection in multiple species**
251 To investigate whether similar functions are selected for on the X in multiple species, we
252 obtained the genes intersecting windows below and above the low diversity threshold using
253 BEDtools v.2.3.0 and the annotation files from the reference genome for each species in our study.
254 Next, we used the gene IDs and matched them to their corresponding ortholog IDs in orthoDB v11
255 (Kuznetsov *et al.* 2023). We then looked for overlapping orthologs across species and obtained the
256 proportion of orthologs that overlap over two, three or four species in low versus high diversity
257 windows.
258 To understand whether there were more shared regions under selection in low versus high
259 diversity regions across species, we obtained 100 random samples of *n* orthologs from high
260 diversity regions with *n* the number of orthologs in low diversity regions per species. From this
261 we computed the mean proportion of shared orthologs under selection as well as the 95%
262 confidence interval in high diversity regions.
263
264 **Code availability**
265 Code used for simulations as well as to process and analyze the data is available at
266 https://github.com/garudlab/DrosCrossSpecies_XchrHardSweeps
267
268 **Results**
269 We analyzed population genomic data of six different *Drosophila* species from seven
270 populations, including: *D. melanogaster* (n=100 Zambia (ZI), n=100 Raleigh (RA)), *D. simulans*
271 (n=170), *D. sechellia* (n=23), *D. mauritiana* (n=15), *D. santomea* (n=7) and *D. teissieri* (n=11)
272 (**Table S2**). To understand not only how selection varies across species, but also across populations
273 within a species, we include a derived *D. melanogaster* population from Raleigh as well as a
274 population from Zambia that is presumed to be within the ancestral range of the species (Pool *et*
275 *al.* 2012). First, we compare the patterns of nucleotide diversity (π) between the autosomes and

276 the X chromosome in the data. Next, we analyze haplotype homozygosity in low diversity regions
277 on the X compared to the rest of the chromosome. Finally, we simulate a variety of models, both
278 without and with selection, to test whether any of these can generate the patterns observed in the
279 data.
280
281 **Patterns of X versus autosomal diversity across the genomes of six *Drosophila* species**
282 When a population is not subject to any selective forces, has the same mutation rate across
283 sexes, a male-to-female ratio of one, and a constant effective population size (N_e), the diversity on
284 the X chromosome (π_X) is expected to be 3/4 of the autosomal diversity (π_A) (Vicoso and
285 Charlesworth 2006). However, such a simple model rarely captures the complex dynamics of
286 natural populations, rather, it is likely that different evolutionary forces affect the X and autosome
287 differently, leading to deviations from the π_X/π_A=0.75 expectation. Across the seven populations
288 studied, we observed a significantly lower diversity on the X compared to autosomes with five out
289 of seven populations showing π_X/π_A below 0.75 (**Fig. 1**). The other two populations (*D.*
290 *melanogaster* (ZI) and *D. teissieri*) also deviated from the π_X/π_A=0.75 expectation but in the
291 opposite direction, with π_X/π_A>0.75. This trend has been previously reported for African
292 populations of *D. melanogaster* and has been attributed to other evolutionary forces such as female
293 sex bias (Kauer *et al.* 2002; Singh *et al.* 2007; Pool *et al.* 2012).
294 Functionally important regions of the genome, such as exons, are expected to be subject to
295 selection, positive or negative (McVicker *et al.* 2009; Arbiza *et al.* 2014; Nam *et al.* 2015). Hence,
296 to investigate the influence of selection across species, we divided the data into three regions:
297 exons, introns, and intergenic regions and computed the ratio of X to autosomal diversity (π_X/π_A)
298 for each region. We found a stronger reduction in diversity in exons compared to introns and
299 intergenic regions for most species (**Fig.1**), pointing to the functional importance of exons (Nam
300 *et al.* 2015). Moreover, we found that π_X/π_A was consistently below 0.75 for all regions for most
301 species (**Fig. 1**), suggesting more directional selection on the X compared to autosomes. However,
302 we noted an exception in *D. sechellia,* where exons exhibited higher levels of diversity than introns
303 and intergenic regions. This behavior could be explained by the high levels of introgression from
304 *D. simulans* to *D. sechellia* (Garrigan *et al.* 2012; Matute and Ayroles 2014; Schrider *et al.* 2018).
305 Introgression tends to be higher in non-functional regions of the genome and is generally higher
306 on the autosomes than on the X due to the involvement of sex chromosomes in hybrid

307 incompatibilities (Muirhead and Presgraves 2016; Turissini and Matute 2017; Fraïsse and
308 Sachdeva 2021). This could lead to a greater difference between X and autosomal diversity in
309 intergenic regions compared to exons, potentially explaining the patterns observed in *D.*
310 *sechellia*'s data.



311
312
313 **Figure 1. Nucleotide diversity (π/bp) in exons, introns and intergenic regions on the**
314 **autosomes and the X chromosome of six *Drosophila* species.** π_X/π_A is lower on exons than on
315 introns and intergenic regions in most species. The red dotted line corresponds to the 0.75
316 expectation under a model with an unbiased sex ration and equal mutation rates on the X and
317 autosomes. For each category we performed 1,000 bootstrap replicates (**Methods**). The tree shown
318 below the X-axis represents the phylogenetic relationship between species as described in FlyBase
319 (Gramates *et al.* 2022).
320
321 To identify potential regions under selection, we calculated π/bp in 20Kb windows across each
322 chromosome for every species. Next, we labeled the analysis windows as low diversity windows

323 if they fell below a diversity threshold set as 20% of the chromosomal π/bp average. Importantly,
324 we computed this threshold for each chromosome separately to normalize for variance in
325 coalescence time across chromosomes. We found that, for all species, the X chromosome showed
326 a higher proportion of windows below the low diversity threshold compared to the autosomes (**Fig.**
327 **2** and **Fig. S4**). Remarkably, these low diversity windows often exhibited π/bp values significantly
328 lower than the defined threshold with values as low 0.5% of the chromosomal average (**Table S3;**
329 **Fig. S5**). Furthermore, we observed multiple instances of consecutive low diversity windows
330 extending up to 200Kb, shown as the yellow shaded regions of **Fig. 2A**. To validate our findings,
331 we repeated this analysis using 10Kb and 50Kb analysis windows as well as two different low
332 diversity thresholds. In all cases, we observed an increased proportion of low diversity windows
333 on the X compared to autosomes (**Fig. S2**).

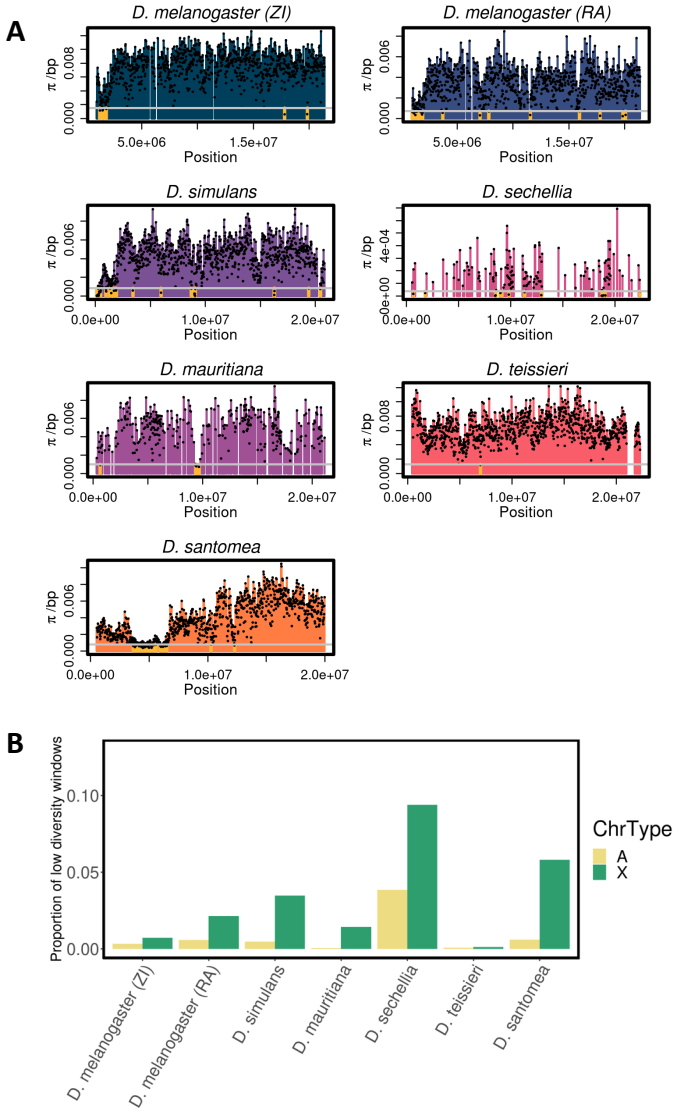
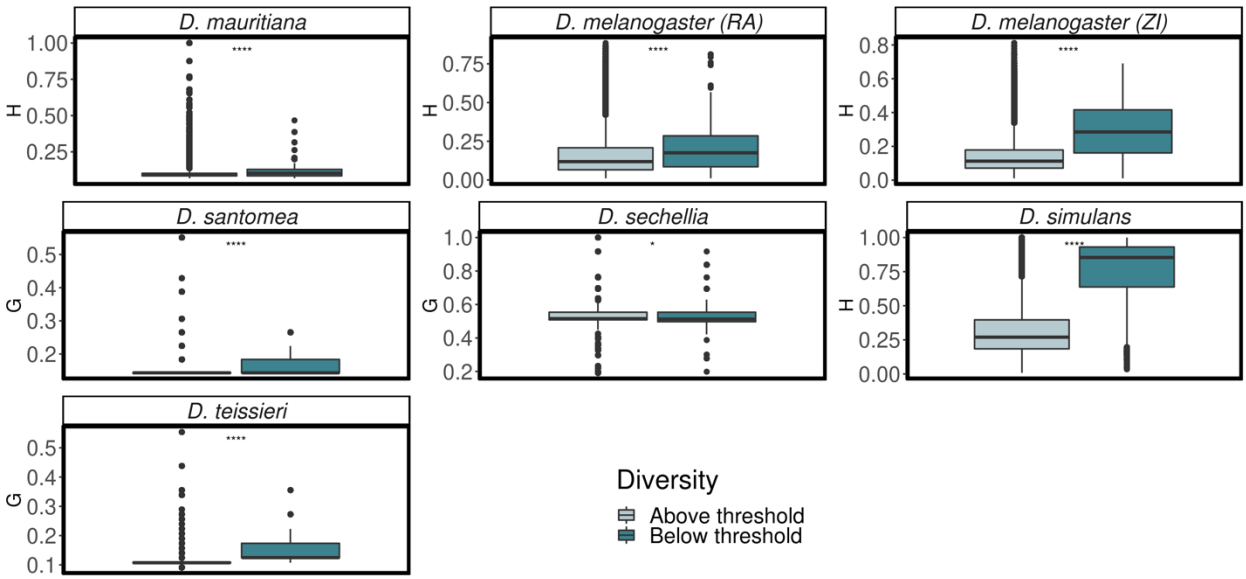


Figure 2. Multiple *Drosophila* species have a higher proportion of windows on the X with steep depressions of diversity compared to autosomes. (A) π/bp in 20kb windows on the X chromosome of six *Drosophila* species and two *D. melanogaster* populations from Raleigh (RA) and Zambia (ZI). Windows where diversity is below 20% of the chromosomal average (gray horizontal line) are highlighted in gold. In **Fig. S4**, 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).

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

345 potential mechanisms responsible for the dips in diversity on the X, we computed haplotype
346 homozygosity (H) across windows below and above the low diversity threshold (**Methods**; **Fig.**
347 **3**) in species with phased genomes (*D. melanogaster, D. simulans* and *D. mauritiana*). For species
348 with unphased genomes, we computed multilocus genotype identity (G) ((Harris *et al.* 2018a);
349 **Methods**). We found that in all populations, windows below the low diversity threshold showed a
350 significantly elevated homozygosity compared to the rest of the chromosome (one-sided Wilcoxon
351 rank sum test p-val< 0.05; **Fig. 3**). Below we examine the role of selection in generating these
352 differential patterns between the X and the autosomes.
353



354
355 **Figure 3. Homozygosity in windows below and above 20% of the average π/bp on the X**
356 **chromosome.** Windows with π/bp below 20% of the chromosomal mean show elevated haplotype
357 homozygosity (H) or elevated multilocus genotype homozygosity (G) compared to windows with
358 π/bp above this threshold (one-sided Wilcoxon rank sum test p-val< 0.05).
359

360 **Neutral models cannot recapitulate diversity patterns observed in the data**

361 The differing diversity patterns on the X versus autosomes could be the result of distinct
362 population genetic forces including: (1) sex biased demography, (2) differences in mutation rates
363 between males and females, (3) lower recombination rates, (4) stronger effects of drift on the X
364 given its smaller N_e and/or (5) more efficient natural selection on the X due to male hemizygosity
365 (Betancourt *et al.* 2004; Vicoso and Charlesworth 2009; Arbiza *et al.* 2014; Nam *et al.* 2015).

366 To gain a better understanding of how different evolutionary processes influence the relative
367 X to autosomal diversity, we quantified π_X, π_A and haplotype homozygosity across a wide variety
368 of scenarios simulated with the population genetics simulator SLiM (Haller and Messer 2019). We
369 started by simulating neutral models of sex bias, low recombination rates, mutation bias and
370 population bottlenecks.
371
372 *Sex bias*
373 Our simulations show that female bias can increase π_X/π_A above the 0.75 expectation (**Fig. 4A**),
374 recapitulating the patterns observed in the Zambian population of *D. melanogaster* and consistent
375 with evidence of female bias in African *D. melanogaster* populations previously reported in the
376 literature (Kauer *et al.* 2002; Dieringer *et al.* 2005; Thornton and Andolfatto 2006; Singh *et al.*
377 2007; Pool *et al.* 2012) **(Fig. 1** and **Fig. 4A)**. *D. teissieri* also has π_X/π_A>0.75, which could also
378 be consistent with female bias. However, none of our female bias models produced elevated
379 haplotype homozygosity compared to neutrality (**Fig. 4B**).
380 Additionally, we found that male sex bias can decrease π_X such that $\pi_X/\pi_A < 0.75$, which could
381 potentially explain low π_X values observed in some species (*D. melanogaster (RA), D. simulans,*
382 *D. mauritiana* and *D. santomea;* **Fig. 1**). However, only the most extreme case of male bias (1:7
383 female-male bias) can marginally increase haplotype homozygosity compared to neutrality (**Fig.**
384 **4B**). To the best of our knowledge, such an extreme male-sex bias in the species studied has not
385 been reported in the literature and therefore seems unlikely.
386
387 *Recombination rate variation*
388 Another possibility is that regions with a low recombination rate produce dips in diversity and
389 high haplotype homozygosity as observed in the data. The footprint of a hard selective sweep is
390 inversely proportional to the recombination rate (~ *s/r*), due to high recombination rates breaking
391 linkage and impeding the formation of long haplotypes at high frequency (Maynard Smith and
392 Haigh 1974; John H. Gillespie 2004). Consequently, it is reasonable to anticipate lower
393 recombination rates in our putative hard sweep regions when compared to the overall genomic
394 landscape. For *D. melanogaster*, where there is a detailed recombination rate available, we find
395 that, on average, recombination rates are 37% and 48% lower in the low diversity regions of the
396 RA and ZI populations, respectively (**Fig. S6**; (Comeron *et al.* 2012). To understand if a lower

397 recombination rate can explain all the signatures in the data, we simulate windows with reduced
398 recombination rates (r_{low}), such that $r_{low}/r = 0.5, 0.2, 0.1, 0.01$. We found that while a low
399 recombination rate can elevate haplotype homozygosity it cannot generate dips of diversity below
400 the low diversity threshold (**Fig. 4, Fig. S7**).
401
402 *Mutation rates*
403 Lower X-linked mutation rates could result in stronger dips in diversity on the X compared to
404 autosomes. A lower X-linked mutation rate is expected when there is a higher number of germline
405 cell divisions in males than females (Drost and Lee 1995; Kirkpatrick and Hall 2004). However,
406 in *Drosophila,* the number of cell divisions in the male and female germlines has been shown to
407 be similar (Drost and Lee 1995), suggesting that there should be no significant difference in the
408 mutation rate between the sexes. Nevertheless, other factors may differentially affect X-linked and
409 autosomal mutation rates such as selection on codon usage on the X given a higher level of codon
410 usage bias and GC content on the X across many *Drosophila* species (Singh *et al.* 2005a; b, 2008;
411 Vicoso *et al.* 2008; Campos *et al.* 2013; Schrider *et al.* 2013; Keightley *et al.* 2014). Past studies
412 on *D. melanogaster* have not shown a statistically significant difference between X-linked and
413 autosomal mutation rates (Keightley *et al.* 2009, 2014; Schrider *et al.* 2013). For other species,
414 evidence of a lower X-linked mutation rate is inconclusive, with few species showing evidence of
415 lower X mutation rates (Garrigan *et al.* 2014), however more studies are needed to better
416 understand X-linked versus autosomal mutation rate differences.
417 Nonetheless, we performed simulations to test the impact of a lower mutation rate on the X
418 and whether this could, on its own, explain the patterns observed in the data. Simulations show
419 that only the most extreme case of lower X-linked mutation rate (μ_X=0.1μ_A) can decrease diversity
420 below the 20% of $\pi_{X,Neutral}$, where $\pi_{X,Neutral}$ is the average π/bp from neutral X chromosome
421 simulations. However, none of our low X-linked mutation rate simulations significantly elevated
422 haplotype homozygosity. In fact, the only case that can considerably reduce $\pi_{X,}$ μ_X=0.1μ_A, shows
423 the lowest haplotype homozygosity across all models (**Fig. 4C**). This is due to longer windows in
424 terms of base pairs for the same SNP window size used across all scenarios which leads to a lower
425 probability of sampling two identical haplotypes and hence a lower haplotype homozygosity. From
426 the above, we conclude that it is unlikely that the low diversity windows on the X are uniquely
427 explained by a lower X-linked mutation rate.

428

429 *Demography*

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

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

448

449

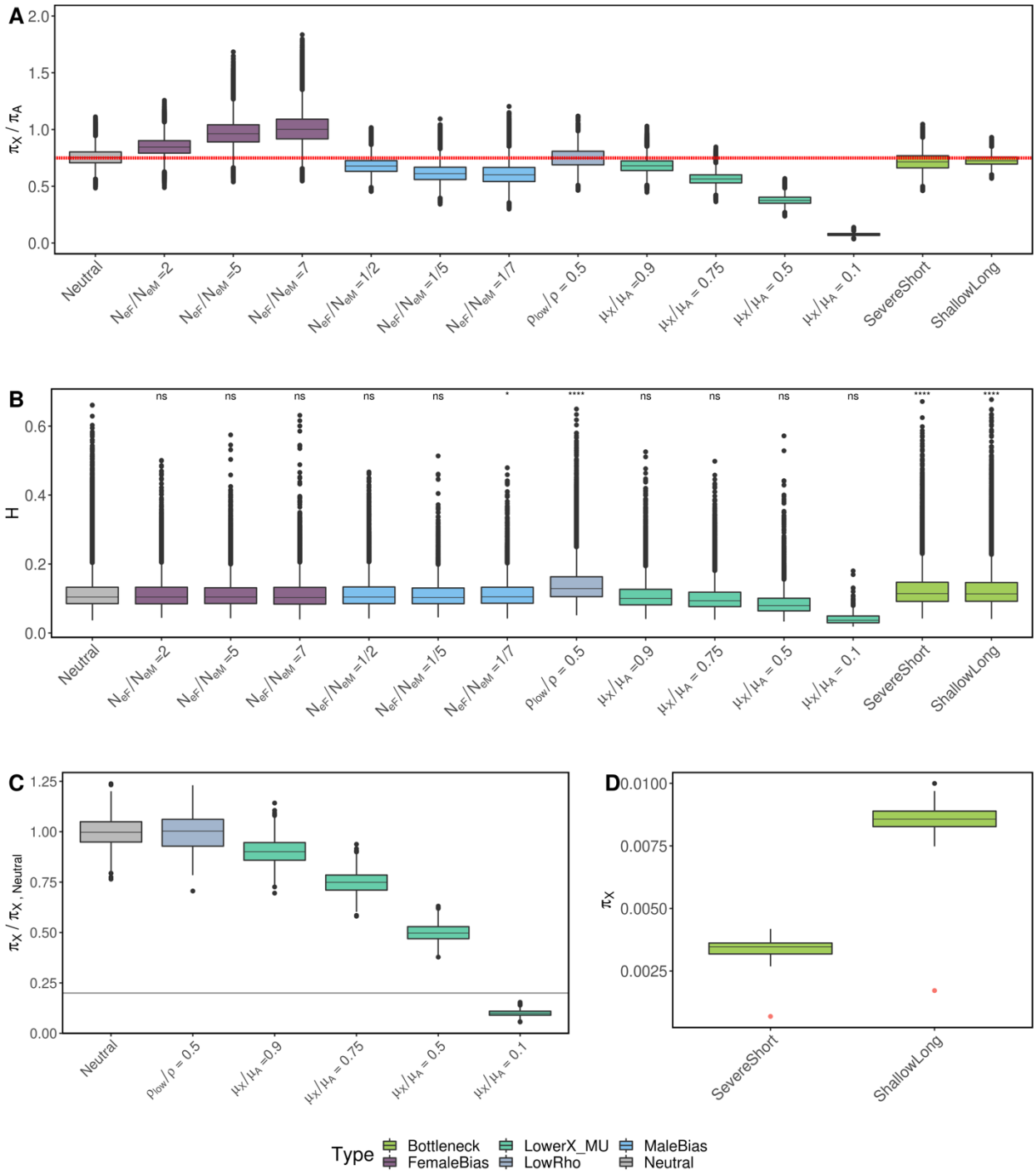


Figure 4. Nucleotide diversity (π) and haplotype homozygosity across 12 models with no selection. The models considered include three female-biased models (N_{eF}/N_{eM}=2, N_{eF}/N_{eM}=5, and N_{eF}/N_{eM}=7), three male-biased models (N_{eM}/N_{eF}=2, N_{eM}/N_{eF} =5, and N_{eM}/N_{eF} = 7), one low recombination model (r_{low}/r = 0.5), four lower X-linked mutation rate models (μ_X/μ_A =0.9, μ_X/μ_A=0.75, μ_X/μ_A=0.5, and μ_X/μ_A=0.1, where μ_X is the X-linked mutation rate and μ_A is the

456 autosomal mutation rate), and two bottleneck models (a severe short bottleneck and a shallow and
457 long bottleneck). **(A)** π_X/π_A across all models, where the red dashed line corresponds to the
458 expected $\pi_X/\pi_A = 0.75$ value in the baseline neutral model (Same sex ratio, equal X and autosomal
459 mutation rates and a constant $N_e=10^6$). **(B)** Haplotype homozygosity across all models. The
460 asterisks represent a significant elevation in haplotype homozygosity relative to complete
461 neutrality using a one-sided Wilcoxon rank-sum test. **(C)** $\pi_X/ \pi_{X,Neutral}$ across low X mutation rate
462 models, with $\pi_{X,Neutral}$ equal to the average π in the baseline neutral model. The horizontal gray line
463 is the low diversity threshold, defined as 20% of the average π/bp of the neutral case. **(D)** π_X in
464 two bottleneck models. The red data point below the boxplots corresponds to 20% of the average
465 π from each bottleneck model.
466
467 **The effect of background selection on nucleotide diversity and haplotype homozygosity**
468 Having investigated neutral evolutionary scenarios, we next examined whether selection
469 can generate the patterns observed in the data. Both background selection (BGS) and selective
470 sweeps can decrease genetic diversity at linked sites, either through the purging of neutral alleles
471 that are linked to deleterious mutations or through hitchhiking, in which a beneficial mutation and
472 its genetic background spread rapidly through the population (Tajima 1989; Charlesworth *et al.*
473 1993, 1995; Stephan 2010). However, background selection is not known to increase haplotype
474 homozygosity (Enard *et al.* 2014; Schrider 2020).
475 To test whether BGS can, on its own, give rise to values of $\pi_X/\pi_A<0.75$, decrease π/bp
476 below the low diversity threshold and elevate haplotype homozygosity, we simulated a population
477 ($Ne=1\text{x}10^6$) in which a fraction (d=0.1,0.5 and 0.8) of mutations are deleterious. The selection
478 coefficient (s_d) for the deleterious mutations followed a gamma distributed DFE with mean and
479 shape parameter -0.000133 and 0.35, respectively (Huber *et al.* 2017). Our simulations show that
480 (i) BGS does not decrease π_X such that π_X/π_A is considerably below 0.75 (**Fig. 5A**) (ii) diversity is
481 reduced below 20% of $\pi_{X,Neutral}$ only when 80% of mutations are deleterious (**Fig. 5B**) and (iii)
482 none of the BGS models considered can elevate homozygosity (**Fig. 5C**).

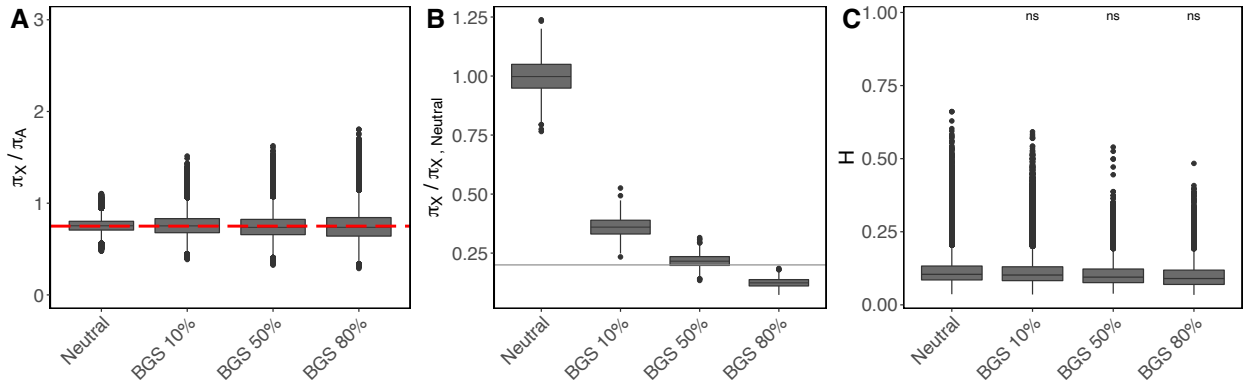


Figure 5. Effect of background selection on diversity and haplotype homozygosity. The models considered include: a neutral scenario with no sex ratio or mutation rate biases and three BGS models. For the BGS models, we varied the proportion of deleterious mutations (10, 50 and 80%) while the DFE for the deleterious selection coefficient (s_d) was gamma distributed with mean and shape parameter -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 one-sided Wilcoxon rank-sum test.

The combined effect of background selection, 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 (BGS, reduced μ_X, low recombination rates, and bottlenecks) cannot produce strong dips in diversity and elevated haplotype 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-7cM/bp, while **Figure S8** shows these results for a lower recombination rate (r=2.5e-7cM/bp). We found

507 that when r = 5e-7cM/bp and 50% or more of new mutations are deleterious, only one of the
508 bottlenecks displayed dips in diversity below the designated low diversity threshold. However,
509 none of these scenarios showed elevated haplotype homozygosity. By contrast, when r = 2.5e-
510 7cM/bp, a single scenario was able to recapitulate all the signatures from the data: a shallow long
511 bottleneck with 50% of deleterious mutations and μ_X/μ_A=0.9. However, empirical evidence of
512 lower X-linked mutation rates in *Drosophila* is inconclusive (see section *Mutation rates* above).
513 Moreover, we note that species with sample sizes of a similar order of magnitude as those of our
514 simulations (*D. melanogaster* and *D. simulans*) exhibit a more pronounced elevation in haplotype
515 homozygosity than what we observe in **Figure S8**. This indicates a larger effect size in our data
516 compared to simulations, where due to the large number of replicates we can detect smaller effect
517 sizes as statistically significant.

518 Finally, to test whether the effect of background selection is stronger when a larger number
519 of deleterious mutations reside on the same chromosome, we simulated longer chromosomes of
520 1Mb instead of 20Kb (**Fig. S9**). These simulations included BGS for three different recombination
521 rates (**Methods**). We found that when recombination is low (r<=1e-8 cM/bp), diversity dips below
522 20% of $\pi_{X,Neutral}$ and haplotype homozygosity is elevated. In this scenario, recombination is not
523 effectively breaking linkage, leading to an elevation of haplotype homozygosity and strong
524 reductions in diversity compared to $\pi_{X,Neutral.}$ However, in this scenario, π_X/π_A does not dip below
525 0.75, which is inconsistent with five of seven populations in the data (**Fig. S9**). Additionally, we
526 note that the fraction of the low diversity windows in the *D. melanogaster* genome that experiences
527 a recombination rate below r=1e-8cM/bp is 0.15, making this unlikely to be the driving force
528 generating dips in diversity.

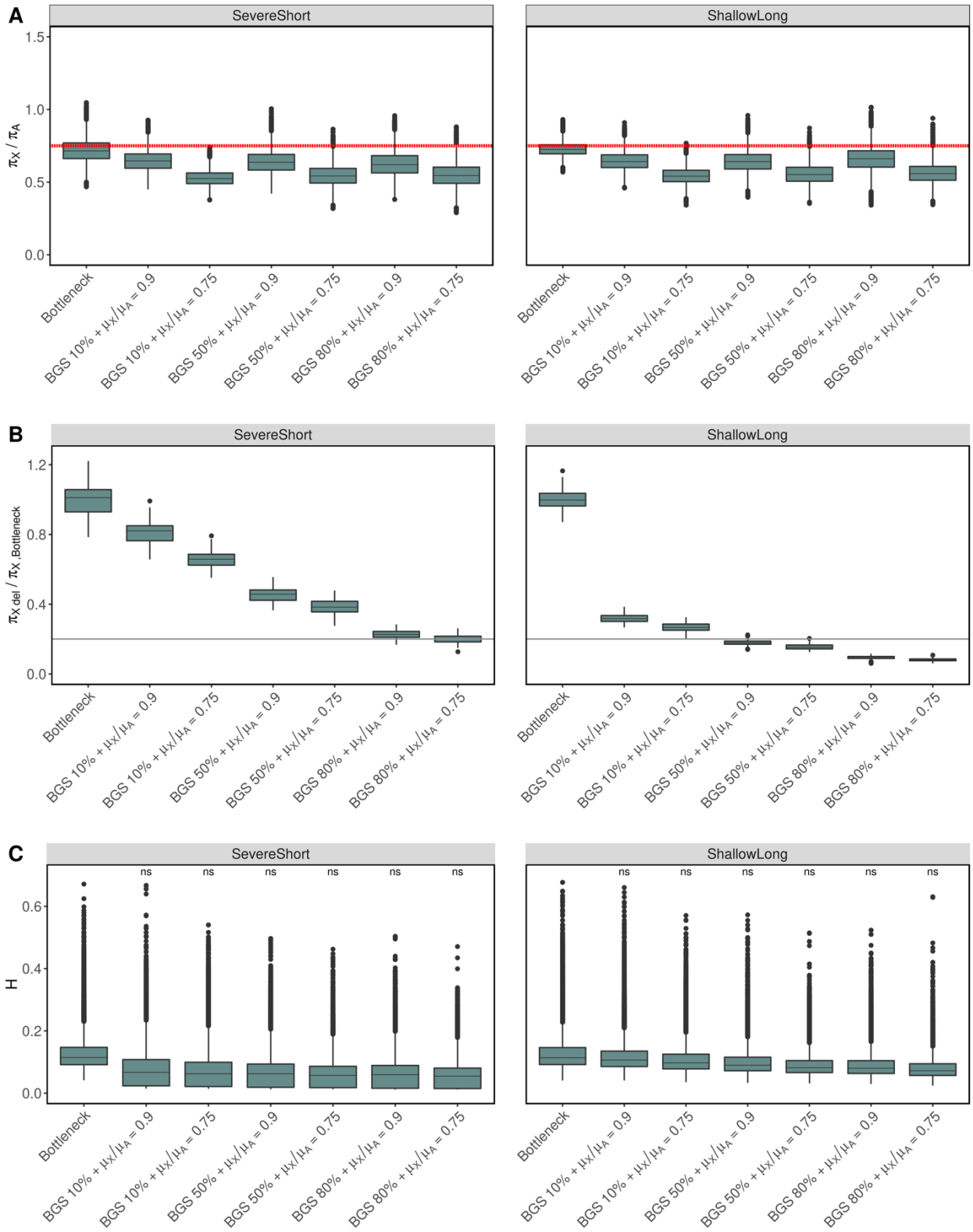


Figure 6. Effect of background selection combined with bottlenecks and low X-linked mutation rate on diversity patterns. The models considered include the combination of a

532 bottleneck model with one of three BGS models each combined with a lower X linked mutation
533 rate (μ_X/μ_A =0.9, μ_X/μ_A=0.75). For the BGS models, we varied the proportion of deleterious
534 mutations (10, 50 and 80%) while the DFE for the deleterious selection coefficient (s_d) was gamma
535 distributed with mean and shape parameter -0.000133 and 0.35, respectively (Huber *et al.* 2017).
536 The recombination rate was set to r=5e-7 cM/bp across all simulations. Figure S8 shows the same
537 corresponding plot for half the recombination rate (2.5e-7 cM/bp). The left column corresponds to
538 a severe short bottleneck model and the right column to a shallow long bottleneck model
539 (**Methods**). For each model, we computed: **(A)** π_X/π_A across all models, where the red dashed line
540 corresponds to the expected π_X/π_A = 0.75 value in the neutral model, **(B)** $\pi_X/\pi_{X,Bottleneck}$ where
541 $\pi_{X,Bottleneck}$ is the average π in the corresponding bottleneck model with no added biases and the
542 solid gray line is 20% of $\pi_{X,Bottleneck}$, **(C)** Haplotype homozygosity across all models and results
543 from a one-sided Wilcoxon rank-sum test for elevation of H in the bottleneck+BGS+Low μ_X
544 scenario compared to a neutral bottleneck model.
545
546 **Hard sweeps can generate the patterns observed on the X across species**
547 Our results indicate that the patterns observed in the data are unlikely to be generated by
548 sex bias, low recombination rates, low X-linked mutations rates, demography or BGS, either
549 individually or in combination. Next, we tested the effect of positive selection in generating dips
550 in diversity and elevated haplotype homozygosity on the X chromosome. To do so, we simulated
551 hard and soft sweeps (**Methods**), varied the strength of selection ($N_e s_b$ = 20, 200 and 2000) and
552 computed π_X/π_A , $\pi_X/\pi_{X,Neutral}$ and haplotype homozygosity in a 20Kb window for each scenario.
553 Our simulations show that selective sweeps can decrease π_X/π_A below 0.75 and elevate
554 haplotype homozygosity (**Fig. 7A, C**), but only hard sweeps can reduce diversity below 20% of
555 $\pi_{X,Neutral}$ (**Fig. 7B**) as long as selection is sufficiently strong (Nes_b =2000). Moreover, compared to
556 the scenarios that we have simulated thus far, we observe a much stronger elevation in haplotype
557 homozygosity when there is positive selection. This is indicative of a more substantial effect size,
558 which aligns better with observations in *D. melanogaster* and *D. simulans,* the species with sample
559 sizes comparable to those from simulations.
560 We note however, that it is also possible for a strong soft sweep that is not too soft ($N_e s_b$
561 $=2000$, θ_a = 0.1 corresponding to approximately two sweeping haplotypes **Fig. S10A**) to reach
562 levels of diversity below the low diversity threshold. However, in this regime (θ_a = 0.1) only

563 slightly more than half (~62%) of the sweeps simulated with $\theta_a = 0.1$ are soft (**Fig. S10**) and,
564 among those that are soft, 48% of the simulations show π_X below the low diversity threshold
565 compared to 82% for the hard sweep model (**Fig. S11**), making hard sweeps a more likely
566 explanation of the data. Thus, we conclude that patterns of reduced diversity and elevated
567 haplotype homozygosity on the X chromosome of the species analyzed are most consistent with
568 hard sweeps.

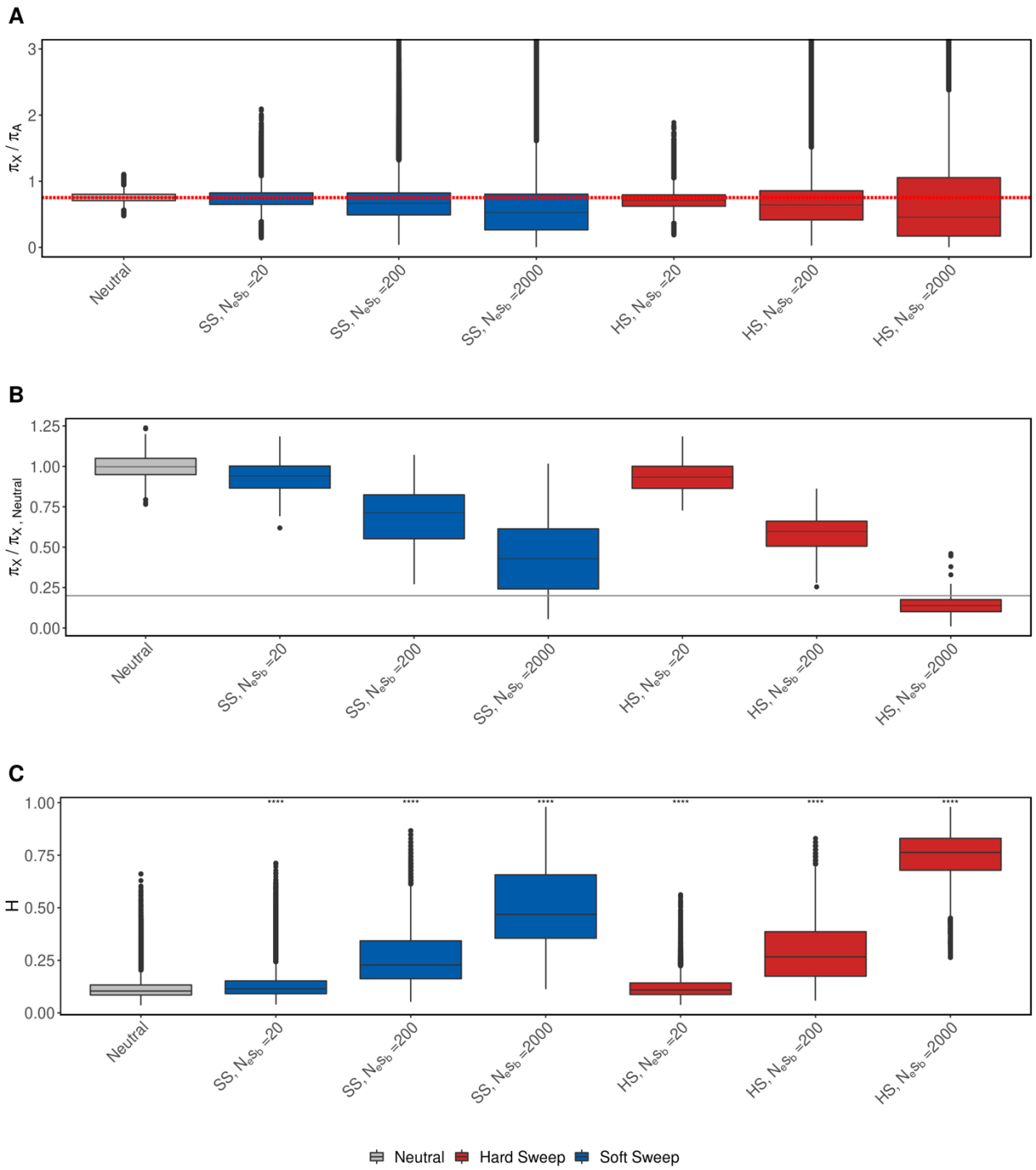


Figure 7. Effect of hard and soft selective sweeps on diversity and haplotype homozygosity. The models considered include: a neutral scenario with no sex ratio or mutation rate biases, three soft sweep (SS) and three hard sweep models (HS). We simulated soft sweep (blue) and hard sweep (red) models (Methods) varying the selection strength of the adaptive mutation ($N_e s_b$=20, 200, 2000). For each model, we computed: **(A)** π_X/π_A, where the red dashed line corresponds to

575 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
576 the π in the baseline neutral model and the solid gray line is 20% of $\pi_{X,Neutral}$; and **(C)** haplotype
577 homozygosity. The asterisks represent a significant elevation in haplotype homozygosity relative
578 to complete neutrality using a one-sided Wilcoxon rank-sum test.
579

580 ## Discussion

581 A classic question in evolutionary biology is how the evolution of the X chromosome
582 differs from that of autosomes given the X chromosome's central role in speciation, brain function,
583 fertility, and sexual dimorphism (Rice 1984; Saifi and Chandra 1999; Skuse 2005; Dean and Mank
584 2014; Payseur *et al.* 2018). Past work has suggested that the X chromosome may exhibit different
585 evolutionary dynamics from the autosomes due to its unique inheritance pattern and increased
586 exposure to selection through male hemizygosity (Vicoso and Charlesworth 2006; Nam *et al.*
587 2015; Charlesworth *et al.* 2018; Muralidhar and Veller 2022; Harris and Garud 2023). Recently,
588 we found evidence in a North American *D. melanogaster* population that the X chromosome
589 experiences an enrichment of hard selective sweeps compared to autosomes due to the increased
590 visibility of new deleterious mutations to natural selection on the hemizygous X of males and
591 reduced effective population size on the X (Harris and Garud 2023). However, it is unclear whether
592 this pattern of enrichment of hard sweeps on the X is a universal feature across all heterogametic
593 species.

594 To understand if the enrichment of hard sweeps on the X is common across species, we
595 analyzed multiple whole genome sequences from six *Drosophila* species and found evidence that
596 suggests that hard sweeps are in fact more common on the X chromosome than autosomes across
597 these species. Additionally, we found that the observed patterns of diversity are inconsistent with
598 background selection, differences in mutation rate, population bottlenecks, or sex bias. We note,
599 however, that when combined, in one specific scenario, low recombination rates, low mutation
600 rates, and BGS could generate both strong dips in diversity and elevated haplotype homozygosity
601 (**Fig. S8**). However, in this scenario we observed a small effect size for the elevation of haplotype
602 homozygosity inconsistent with the data (**Fig. 3**, **Fig. S8**). Additionally, we note that there has not
603 been any definitive evidence of reduced mutation rates on the X relative to autosomes in
604 *Drosophila*.

605 Furthermore, we found that soft sweeps generally cannot generate the patterns observed in
606 the data, except for when soft sweeps are not too soft (e.g. there are only 2 sweeping haplotypes).
607 However, in this scenario (θ_a =0.1), only 62% of the simulations led to soft sweeps, and of those,
608 only 48% generated dips in diversity falling below the low diversity threshold, whereas 82% of
609 hard sweeps can. Therefore, we acknowledge that it is possible that other factors other than hard
610 sweeps could be responsible for generating some of the low diversity windows observed in the
611 data. However, our results suggest that these forces are unlikely to be the dominant processes
612 driving the patterns in the data. Future work will be important for disentangling the effects of these
613 forcers on diversity on the X.

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

626 Comparative population genetics, in which multiple genomes from several species each
627 are considered simultaneously, is also a relatively new area of work given the paucity of deep
628 population genetic samples from multiple species. Population genetic studies have traditionally
629 focused on analyzing multiple genomes from a single species, e.g. (Arbiza *et al.* 2014; Garud and
630 Rosenberg 2015; Signor *et al.* 2018) with few examining more than two population genetic whole
631 genome datasets from different species simultaneously ((Nam *et al.* 2015; Chen *et al.* 2018;
632 Nadachowska-Brzyska *et al.* 2019; Latrille *et al.* 2023; Rodrigues *et al.* 2023) with (Latrille *et al.*
633 2023) restricted to exomes), leaving open numerous avenues of inquiry on the commonalities and
634 idiosyncrasies of population genetic processes.

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

654 Despite individual species showing deviations from the trend, some species exhibited
655 commonalities. For example, the species analyzed in this study all had a higher proportion of low
656 diversity regions on the X chromosome compared to the autosomes. Additionally, we found some
657 targets of selection to be shared across species, however they were generally few in number (**Fig.**
658 **S12-S14**, **Text S1**). Instead, most genes found in low diversity windows were found in a single
659 species only, indicating that while hard sweeps might be common on the X, different functions
660 may be under selection in different species. We note that this finding is distinct from that of
661 primates (Nam *et al.* 2015) where there appears to be more overlap across species. This suggests
662 that the underlying mechanisms of selection in *Drosophila* versus primates may differ.

663 Our paper is a clear demonstration of how comparative population genetic analyses can
664 reveal new insights into the forces that shape genetic variation within and across species, which
665 we expect to serve as a useful example as new high-resolution population genomic datasets from

666 multiple species become increasingly available in the coming years. While our study did not
667 examine the prevalence of soft sweeps in autosomes and the X chromosome due to limited samples
668 sizes per species, future research with larger sample sizes could provide the relevant data needed
669 to be able to detect soft sweeps with haplotype homozygosity statistics, thereby providing a more
670 comprehensive understanding of the tempo and mode of adaptation across species (Pennings and
671 Hermisson 2006b). Ultimately, our study highlights the significance of hard sweeps in shaping the
672 diversity patterns of the X chromosome across species and suggests an important evolutionary
673 mechanism that may be widespread among all species. Future work may reveal the potential role
674 of these hard sweeps in driving sexual dimorphism and speciation given the X chromosome's
675 significant involvement in these important processes.

676 Data availability

677 Sequence data for *D. melanogaster* can be downloaded from www.johnpool.net . For *D. simulans*
678 the vcf file can be downloaded from https://zenodo.org/record/154261#.YzMzty2z3jC
679 For the remaining species, sequence data are available at NCBI's Short Read Archive with
680 accession numbers given in Table S1.
681

682 Acknowledgments

683 We are grateful to Rebekah Rogers and Peter Andolfatto for providing the *D.simulans* w501
684 genome assembly. We thank Dmitri Petrov and Kirk Lohmueller for helpful conversations. MH
685 was supported by the National Institutes of Health Systems and Integrative Biology Training Grant
686 (NIH-NIGMS 5T32GM008185) and the National Institutes of Health Training Grant in Genomic
687 Analysis and Interpretation (NIH T32HG002536). BYK was supported by NIH grant NIGMS
688 F32GM135998. NRG was supported by funding from a National Science Foundation CAREER
689 award.
690

691 References

692 Arbiza L., S. Gottipati, A. Siepel, and A. Keinan, 2014 Contrasting X-linked and autosomal
693 diversity across 14 human populations. Am J Hum Genet 94: 827–844.
694 https://doi.org/10.1016/j.ajhg.2014.04.011
695 Assaf Z. J., D. A. Petrov, and J. R. Blundell, 2015 Obstruction of adaptation in diploids by
696 recessive, strongly deleterious alleles. Proc Natl Acad Sci U S A 112: E2658–E2666.
697 https://doi.org/10.1073/pnas.1424949112

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

Betancourt A. J., Y. Kim, and H. A. Orr, 2004 A pseudohitchhiking model of X vs. autosomal diversity. Genetics 168: 2261–2269. https://doi.org/10.1534/genetics.104.030999

Campos J. L., K. Zeng, D. J. Parker, B. Charlesworth, and P. R. Haddrill, 2013 Codon usage bias and effective population sizes on the X chromosome versus the autosomes in drosophila melanogaster. Mol Biol Evol 30: 811–823. https://doi.org/10.1093/molbev/mss222

Charlesworth B., J. A. Coyne, and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. American Naturalist 130: 113–146. https://doi.org/10.1086/284701

Charlesworth B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

Charlesworth D., B. Charlesworth, and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics 141: 1619–1632. https://doi.org/10.1093/genetics/141.4.1619

Charlesworth B., J. L. Campos, and B. C. Jackson, 2018 Faster-X evolution: Theory and evidence from Drosophila. Mol Ecol 27: 3753–3771. https://doi.org/10.1111/mec.14534

Chen Z. H., M. Zhang, F. H. Lv, X. Ren, W. R. Li, *et al.*, 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: 1282–1297. https://doi.org/10.1093/gbe/evy085

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

Comeron J. M., R. Ratnappan, and S. Bailin, 2012 The Many Landscapes of Recombination in Drosophila melanogaster. PLoS Genet 8. https://doi.org/10.1371/journal.pgen.1002905

Coyne J. A., 1998 The evolutionary genetics of speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 353: 287–305. https://doi.org/10.1098/rstb.1998.0210

Dean R., and J. E. Mank, 2014 The role of sex chromosomes in sexual dimorphism: Discordance between molecular and phenotypic data. J Evol Biol 27: 1443–1453. https://doi.org/10.1111/jeb.12345

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

Drost J. B., and W. R. Lee, 1995 Biological basis of germline mutation: Comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. Environ Mol Mutagen 25: 48–64. https://doi.org/10.1002/em.2850250609

Enard D., P. W. Messer, and D. A. Petrov, 2014 Genome-wide signals of positive selection in human evolution. Genome Res 24: 885–895. https://doi.org/10.1101/gr.164822.113

Feder A. F., P. S. Pennings, and D. A. Petrov, 2021 The clarifying role of time series data in the population genetics of HIV. PLoS Genet 17.

Fraïsse C., and H. Sachdeva, 2021 The rates of introgression and barriers to genetic exchange between hybridizing species: Sex chromosomes vs autosomes. Genetics 217. https://doi.org/10.1093/GENETICS/IYAA025

Garrigan D., S. B. Kingan, A. J. Geneva, P. Andolfatto, A. G. Clark, *et al.*, 2012 Genome sequencing reveals complex speciation in the Drosophila simulans clade. Genome Res 22: 1499–1511. https://doi.org/10.1101/gr.130922.111

Garrigan D., S. B. Kingan, A. J. Geneva, J. P. Vedanayagam, and D. C. Presgraves, 2014 Genome diversity and divergence in Drosophila mauritiana: Multiple signatures of faster X evolution. Genome Biol Evol 6: 2444–2458. https://doi.org/10.1093/gbe/evu198

Garud N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. PLoS Genet 11: 1–32. https://doi.org/10.1371/journal.pgen.1005004

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

Garud N. R., P. W. Messer, and D. A. Petrov, 2021 Detection of hard and soft selective sweeps from Drosophila melanogaster population genomic data. PLoS Genet 17. https://doi.org/10.1371/journal.pgen.1009373

Glaser-Schmitt A., M. J. Wittmann, T. J. S. Ramnarine, and J. Parsch, 2021 Sexual Antagonism, Temporally Fluctuating Selection, and Variable Dominance Affect a Regulatory Polymorphism in Drosophila melanogaster . Mol Biol Evol 38: 4891–4907. https://doi.org/10.1093/molbev/msab215

Gramates L. S., J. Agapite, H. Attrill, B. R. Calvi, M. A. Crosby, *et al.*, 2022 FlyBase: a guided tour of highlighted features. Genetics 220. https://doi.org/10.1093/genetics/iyac035

Haller B. C., and P. W. Messer, 2019 SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. Mol Biol Evol 36: 632–637. https://doi.org/10.1093/molbev/msy228

Harris A. M., N. R. Garud, and M. Degiorgio, 2018a Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. Genetics 210: 1429–1452. https://doi.org/10.1534/genetics.118.301502

Harris R. B., A. Sackman, and J. D. Jensen, 2018b On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. PLoS Genet 14: e1007859. https://doi.org/10.1371/JOURNAL.PGEN.1007859

Harris M., and N. R. Garud, 2023 Enrichment of Hard Sweeps on the X Chromosome in Drosophila melanogaster. Mol Biol Evol 40: 1–16. https://doi.org/10.1093/molbev/msac268

Hermisson J., and P. S. Pennings, 2005 Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics 169: 2335–2352. https://doi.org/10.1534/genetics.104.036947

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

Huber C. D., B. Y. Kim, C. D. Marsden, and K. E. Lohmueller, 2017 Determining the factors driving selective effects of new nonsynonymous mutations. Proc Natl Acad Sci U S A 114: 4465–4470. https://doi.org/10.1073/pnas.1619508114

John H. Gillespie, 2004 *Population Genetics: A Concise Guide*. Johns Hopkins University Press.

Johri P., B. Charlesworth, and J. D. Jensen, 2020 Toward an evolutionarily appropriate null model: Jointly inferring demography and purifying selection. Genetics 215: 173–192. https://doi.org/10.1534/genetics.119.303002

Kaplan N. L., R. R. Hudson, and C. H. Langley, 1989 The “hitchhiking effect” revisited. Genetics 123: 887–899. https://doi.org/10.1093/genetics/123.4.887

Kauer M., B. Zangerl, D. Dieringer, and C. Schlötterer, 2002 Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of Drosophila melanogaster. Genetics 160: 247–256.

Keightley P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar, *et al.*, 2009 Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. Genome Res 19: 1195–1201. https://doi.org/10.1101/gr.091231.109

Keightley P. D., R. W. Ness, D. L. Halligan, and P. R. Haddrill, 2014 Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. Genetics 196: 313–320. https://doi.org/10.1534/genetics.113.158758

Kirkpatrick M., and D. W. Hall, 2004 *Male-biased mutation, sex linkage, and the rate of adaptive evolution*.

Kuznetsov D., F. Tegenfeldt, M. Manni, M. Seppey, M. Berkeley, *et al.*, 2023 OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res 51: D445–D451. https://doi.org/10.1093/nar/gkac998

Lack J. B., C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig, *et al.*, 2015 The drosophila genome nexus: A population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. Genetics 199: 1229–1241. https://doi.org/10.1534/genetics.115.174664

Latrille T., N. Rodrigue, and N. Lartillot, 2023 Genes and sites under adaptation at the phylogenetic scale also exhibit adaptation at the population-genetic scale. Proceedings of the National Academy of Sciences 120. https://doi.org/10.1073/pnas.2214977120

Lawrie D. S., P. W. Messer, R. Hershberg, and D. A. Petrov, 2013 Strong Purifying Selection at Synonymous Sites in D. melanogaster. PLoS Genet 9. https://doi.org/10.1371/journal.pgen.1003527

Legrand D., M. I. Tenaillon, P. Matyot, J. Gerlach, D. Lachaise, *et al.*, 2009 Species-wide genetic variation and demographic history of Drosophila sechellia, a species lacking population structure. Genetics 182: 1197–1206. https://doi.org/10.1534/genetics.108.092080

Levine R. P., 1955 Chromosome Structure and the Mechanism of Crossing Over. Proceedings of the National Academy of Sciences 41: 727–730. https://doi.org/10.1073/pnas.41.10.727

Li H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv.

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

Maroja L. S., E. L. Larson, S. M. Bogdanowicz, and R. G. Harrison, 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: Genes, Genomes, Genetics 5: 2219–2227. https://doi.org/10.1534/g3.115.021246

Mather K., 1939 Crossing Over and Heterochromatin in the X Chromosome of Drosophila Melanogaster. Genetics 24: 413–435. https://doi.org/10.1093/genetics/24.3.413

Matute D. R., and J. F. Ayroles, 2014 Hybridization occurs between Drosophila simulans and D. sechellia in the Seychelles archipelago. J Evol Biol 27: 1057–1068. https://doi.org/10.1111/jeb.12391

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

McGrath C., 2022 Highlight: Comparative Population Genomics—Answering Old Questions with New Data. Genome Biol Evol 14.

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

Meany M. K., W. R. Conner, S. V. Richter, J. A. Bailey, M. Turelli, *et al.*, 2019 Loss of cytoplasmic incompatibility and minimal fecundity effects explain relatively low Wolbachia frequencies in Drosophila mauritiana. Evolution (N Y) 73: 1278–1295. https://doi.org/10.1111/evo.13745

Messer P. W., and D. A. Petrov, 2013 Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol 28: 659–669. https://doi.org/10.1016/j.tree.2013.08.003

Muirhead C. A., and D. C. Presgraves, 2016 Hybrid incompatibilities, local adaptation, and the genomic distribution of natural introgression between species. American Naturalist 187: 249–261. https://doi.org/10.1086/684583

Muralidhar P., and C. Veller, 2022 Dominance shifts increase the likelihood of soft selective sweeps. Evolution (N Y) 1–19. https://doi.org/10.1111/evo.14459

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

Nam K., K. Munch, A. Hobolth, J. Y. Dutheil, K. R. Veeramah, *et al.*, 2015 Extreme selective sweeps independently targeted the X chromosomes of the great apes. Proc Natl Acad Sci U S A 112: 6413–6418. https://doi.org/10.1073/pnas.1419306112

O'Leary N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, *et al.*, 2016 Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44: D733–D745. https://doi.org/10.1093/nar/gkv1189

Orr H. A., and A. J. Betancourt, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics 157: 875–884. https://doi.org/10.1093/genetics/157.2.875

Patten M. M., 2019 The X chromosome favors males under sexually antagonistic selection. Evolution (N Y) 73: 84–91. https://doi.org/10.1111/evo.13646

Payseur B. A., D. C. Presgraves, and D. A. Filatov, 2018 Introduction: Sex chromosomes and speciation, pp. 3745–3748 in *Molecular Ecology*,.

Pennings P. S., and J. Hermisson, 2006a Soft sweeps II - Molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol 23: 1076–1084. https://doi.org/10.1093/molbev/msj117

Pennings P. S., and J. Hermisson, 2006b Soft sweeps III: The signature of positive selection from recurrent mutation. PLoS Genet 2: 1998–2012. https://doi.org/10.1371/journal.pgen.0020186

Peter B. M., E. Huerta-Sanchez, and R. Nielsen, 2012 Distinguishing between Selective Sweeps from Standing Variation and from a De Novo Mutation. PLoS Genet 8: e1003011. https://doi.org/10.1371/journal.pgen.1003011

Pool J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno, *et al.*, 2012 Population Genomics of Sub-Saharan Drosophila melanogaster: African Diversity and Non-African Admixture. PLoS Genet 8. https://doi.org/10.1371/journal.pgen.1003080

Quinlan A. R., and I. M. Hall, 2010 BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ragsdale A. P., and R. N. Gutenkunst, 2017 Inferring demographic history using two-locus statistics. Genetics 206: 1037–1048. https://doi.org/10.1534/genetics.117.201251

Rice W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution (N Y) 38: 735–742. https://doi.org/10.1111/j.1558-5646.1984.tb00346.x

Rodrigues M. F., A. D. Kern, and P. L. Ralph, 2023 Shared evolutionary processes shape landscapes of genomic variation in the great apes. bioRxiv.

Rogers R. L., L. Shao, J. S. Sanjak, P. Andolfatto, and K. R. Thornton, 2014 Revised Annotations, sex-biased expression, and lineage-specific genes in the Drosophila melanogaster group. G3: Genes, Genomes, Genetics 4: 2345–2351. https://doi.org/10.1534/g3.114.013532

Saifi G. M., and H. S. Chandra, 1999 An apparent excess of sex and reproduction-related genes on the human X chromosome. Proceedings of the Royal Society B: Biological Sciences 266: 203–209. https://doi.org/10.1098/rspb.1999.0623

Sankararaman S., S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, *et al.*, 2014 The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507: 354–357. https://doi.org/10.1038/nature12961

Schrider D. R., D. Houle, M. Lynch, and M. W. Hahn, 2013 Rates and genomic consequences of spontaneous mutational events in Drosophila melanogaster. Genetics 194: 937–954. https://doi.org/10.1534/genetics.113.151670

Schrider D. R., F. K. Mendes, M. W. Hahn, and A. D. Kern, 2015 Soft shoulders ahead: Spurious signatures of soft and partial selective sweeps result from linked hard sweeps. Genetics 200: 267–284. https://doi.org/10.1534/genetics.115.174912

Schrider D. R., and A. D. Kern, 2016 S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning. PLoS Genet 12. https://doi.org/10.1371/journal.pgen.1005928

Schrider D. R., J. Ayroles, D. R. Matute, and A. D. Kern, 2018 Supervised machine learning reveals introgressed loci in the genomes of Drosophila simulans and D. Sechellia. PLoS Genet 14. https://doi.org/10.1371/JOURNAL.PGEN.1007341

Schrider D. R., 2020 Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. Genetics 216: 499–519. https://doi.org/10.1534/genetics.120.303469

Serrato-Capuchina A., E. R. R. D'Agostino, D. Peede, B. Roy, K. Isbell, *et al.*, 2021 P-elements strengthen reproductive isolation within the Drosophila simulans species complex. Evolution (N Y) 75: 2425–2440. https://doi.org/10.1111/evo.14319

Sheehan S., and Y. S. Song, 2016 Deep Learning for Population Genetic Inference. PLoS Comput Biol 12: e1004845. https://doi.org/10.1371/journal.pcbi.1004845

Signor S. A., F. N. New, and S. Nuzhdin, 2018 A large panel of drosophila simulans reveals an abundance of common variants. Genome Biol Evol 10: 189–206. https://doi.org/10.1093/gbe/evx262

Singh N. D., J. C. Davis, and D. A. Petrov, 2005a X-linked genes evolve higher codon bias in Drosophila and Caenorhabditis. Genetics 171: 145–155. https://doi.org/10.1534/genetics.105.043497

Singh N. D., J. C. Davis, and D. A. Petrov, 2005b Codon bias and noncoding GC content correlate negatively with recombination rate on the Drosophila X chromosome. J Mol Evol 61: 315–324. https://doi.org/10.1007/s00239-004-0287-1

Singh N. D., J. M. Macpherson, J. D. Jensen, and D. A. Petrov, 2007 Similar levels of X-linked and autosomal nucleotide variation in African and non-African populations of Drosophila melanogaster. BMC Evol Biol 7: 1–16. https://doi.org/10.1186/1471-2148-7-202

Singh N. D., A. M. Larracuente, and A. G. Clark, 2008 Contrasting the efficacy of selection on the X and autosomes in drosophila. Mol Biol Evol 25: 454–467. https://doi.org/10.1093/molbev/msm275

Skuse D. H., 2005 X-linked genes and mental functioning. Hum Mol Genet 14: 27–32. https://doi.org/10.1093/hmg/ddi112

Smit A. F. A., R. Hubley, and P. Grenn, 2015 RepeatMasker Open-4.0. RepeatMasker Open.

Stephan W., 2010 Genetic hitchhiking versus background selection: The controversy and its implications. Philosophical Transactions of the Royal Society B: Biological Sciences 365: 1245–1253.

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

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

Turelli M., and H. A. Orr, 1995 The dominance theory of HALDANE's rule. Genetics 140: 389–402. https://doi.org/10.1093/genetics/140.1.389

Turissini D. A., and D. R. Matute, 2017 Fine scale mapping of genomic introgressions within the Drosophila yakuba clade. PLoS Genet 13. https://doi.org/10.1371/journal.pgen.1006971

Vicoso B., and B. Charlesworth, 2006 Evolution on the X chromosome: Unusual patterns and processes. Nat Rev Genet 7: 645–653.

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

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

Vincenten N., L. M. Kuhl, I. Lam, A. Oke, A. R. W. Kerr, *et al.*, 2015 The kinetochore prevents centromere-proximal crossover recombination during meiosis. Elife 4. https://doi.org/10.7554/eLife.10850

Wall J. D., and J. K. Pritchard, 2003 Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet 4: 587–597.

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

Mariana Harris, Bernard Kim, and Nandita Garud

Table of Contents

Supplemental Tables:

Table S1	*D.mauritiana, D. teissieri, D. sechellia* and *D. santomea* sequences.	Page 3
Table S2	Drosophila species and populations used in study	Page 7
Table S3	Minimum and mean π/bp computed in non-overlapping 20Kb windows on the X chromosome across species	Page 8

Supplemental Figures:

Figure S1	Phred-scaled quality score (QUAL) across species.	Page 9
Figure S2	Low diversity windows on the X versus autosomes as a function of the low diversity thresholds and window size.	Page 10
Figure S3	Proportion of missing data and Phred-scaled quality score (QUAL) in windows above and below the low diversity threshold for the X chromosome.	Page 11
Figure S4	Genome-wide diversity in six species.	Page 12
Figure S5	Distribution of π/bp for the X chromosome across species.	Page 13
Figure S6	*D. melanogaster* recombination rates (cM/bp) in regions below and above the low diversity threshold.	Page 14
Figure S7	Windows of low recombination can elevate haplotype homozygosity but cannot reduce diversity below 20% of $\pi_{neutral}$.	Page 15
Figure S8	Background selection combined with bottlenecks and low X-linked mutation rate for r= 2.5e-7cM/bp.	Page 16
Figure S9	Nucleotide diversity (π) and haplotype homozygosity (H) in 1Mb simulated chromosomes with BGS.	Page 18
Figure S10	Number of origins and proportion of hard/soft sweep simulations as a function of θa	Page 20

Figure S11	Diversity (π) and haplotype homozygosity (H) in hard and soft sweeps	Page 21
Figure S12	Proportion of orthologs shared across species in windows below and above the low diversity threshold	Page 23
Figure S13	Proportion of orthologs shared across species in windows below and above the low diversity threshold in *D.melanogaster* (ZI) and *D. melanogaster* (RA) only.	Page 24
Figure S14	Number of orthologs shared across species.	Page 25

Supplemental Texts:

Text S1	Overlapping orthologous gene groups with low diversity on the X chromosome across species	Page 26

Table S1. *D.mauritiana, D. teissieri, D. sechellia* and *D. santomea* sequences.

Species	Sample ID	BioProject	Accession	Ploidy	Read length	Bases (bp)	Geographic location	Citation
D. mauritiana	Dmau_R23	PRJNA158675	SRR1560090	1	76	6573297456	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R31	PRJNA158675	SRR1560098	1	76	6615850008	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R32	PRJNA158675	SRR1560103	1	76	6764876128	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R39	PRJNA158675	SRR1560110	1	76	6997261176	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R41	PRJNA158675	SRR1560132	1	76	6518381680	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R44	PRJNA158675	SRR1560147	1	76	6882623992	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R56	PRJNA158675	SRR1560150	1	76	6840838888	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R61	PRJNA158675	SRR1560268	1	76	6421443832	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R8	PRJNA158675	SRR1560276	1	76	6374197216	Mauritius: Le Reduit	Garrigan et al. 2014 GBE
D. mauritiana	Dmau_R50	PRJNA719437	SRR14138479	1	150	8276949900	Mauritius (?)	Serrato-Capuchina et al. 2021 Evolution
D. mauritiana	Dmau_R42	PRJNA719437	SRR14138484	1	150	6624496500	Mauritius	Serrato-Capuchina et al. 2021 Evolution
D. mauritiana	Dmau_R60	PRJNA530272	SRR8834567	1	150	4640534100	Mauritius	Meany et al. 2019 Evolution
D. mauritiana	Dmau_R9	PRJNA530272	SRR8834568	1	150	5240508300	Mauritius	Meany et al. 2019 Evolution
D. mauritiana	Dmau_R29	PRJNA530272	SRR8834569	1	150	5050271700	Mauritius	Meany et al. 2019 Evolution
D. mauritiana	Dmau_mau12	PRJNA518878	SRR10728582	1	150	8722362900	NA	Sproul et al. 2020 MBE
D. teissieri	Dtei_La_Lope_Gabon	PRJNA395473	SRR5860571	1	150	5893400700	Gabon	Turissini et al. 2017 Plos Gen

D. teissieri	Dtei_cascade_4_2	PRJNA395473	SRR5860572	1	150	6481090200	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_Balancha_1	PRJNA395473	SRR5860615	1	150	4843188900	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_cascade_2_2	PRJNA395473	SRR5860616	1	150	5186208900	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_cascade_4_3	PRJNA395473	SRR5860617	1	150	3906514500	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_cascade_4_1	PRJNA395473	SRR5860618	1	150	4406747700	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_Zimbabwe	PRJNA395473	SRR5860619	1	150	5200208400	Zimbabwe	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_Selinda	PRJNA395473	SRR5860620	1	150	4592531700	Zimbabwe	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_House_Bioko	PRJNA395473	SRR5860621	1	150	5739754800	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_cascade_2_4	PRJNA395473	SRR5860622	1	150	4534532100	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. teissieri	Dtei_cascade_2_1	PRJNA395473	SRR5860623	1	150	4674675300	Equatorial Guinea: Bioko	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_maria_3	PRJNA395473	SRR5860570	2	150	6114171000	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_Anro_B8	PRJNA395473	SRR5860573	2	150	5269412100	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_LD8	PRJNA395473	SRR5860582	2	150	6028197600	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_PNF4	PRJNA395473	SRR5860583	2	150	4647707100	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_LD15	PRJNA395473	SRR5860584	2	150	6799504800	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_Anro_B7	PRJNA395473	SRR5860645	2	150	6055820700	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_LD16	PRJNA395473	SRR5860656	2	150	5622810000	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_LD13	PRJNA395473	SRR5860657	2	150	6386079900	Seychelles	Turissini et al. 2017 Plos Gen

D. sechellia	Dsech_LD14	PRJNA395473	SRR5860658	2	150	4884200700	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_mariane_1	PRJNA395473	SRR5860659	2	150	7521924900	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_Anro_B1	PRJNA395473	SRR5860660	2	150	5327670900	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_Anro_B2	PRJNA395473	SRR5860661	2	150	4983445200	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_LD12	PRJNA395473	SRR5860662	2	150	5088897000	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_Anro_B6	PRJNA395473	SRR5860663	2	150	5065029600	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_LD11_sech	PRJNA395473	SRR5860664	2	150	6451788300	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_Anro_B3	PRJNA395473	SRR5860665	2	150	6008176800	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_PNF10	PRJNA395473	SRR5860666	2	150	4788343500	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_PNF11	PRJNA395473	SRR5860667	2	150	5409410700	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_PNF8	PRJNA395473	SRR5860672	2	150	7880448600	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_PNF5	PRJNA395473	SRR5860673	2	150	6064632600	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_PNF3	PRJNA395473	SRR5860674	2	150	6464577600	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_PNF7	PRJNA395473	SRR5860675	2	150	4281983100	Seychelles	Turissini et al. 2017 Plos Gen
D. sechellia	Dsech_Anro_B5	PRJNA395473	SRR5860676	2	150	5997454500	Seychelles	Turissini et al. 2017 Plos Gen
D. santomea	Dsan_C650_14	PRJNA395473	SRR5860635	2	125	9303779000	Sao Tome and Principe: Sao Tome	Turissini et al. 2017 Plos Gen
D. santomea	Dsan_Rain42	PRJNA395473	SRR5860636	2	125	10322705500	Sao Tome and Principe: Sao Tome	Turissini et al. 2017 Plos Gen

D. santomea	Dsan_CAR1600	PRJNA395473	SRR5860637	2	125	10440031000	Sao Tome and Principe: Sao Tome	Turissini et al. 2017 Plos Gen
D. santomea	Dsan_san_Field3	PRJNA395473	SRR5860638	2	125	10190783000	Sao Tome and Principe: Sao Tome	Turissini et al. 2017 Plos Gen
D. santomea	Dsan_BS14	PRJNA395473	SRR5860641	2	125	15530315250	Sao Tome and Principe: Sao Tome	Turissini et al. 2017 Plos Gen
D. santomea	Dsan_C550_39	PRJNA395473	SRR5860642	2	125	10438707750	Sao Tome and Principe: Sao Tome	Turissini et al. 2017 Plos Gen
D. santomea	Dsan_STOCAGO_1482	PRJNA647033	SRR12340817	2	150	15900107928	Sao Tome and Principe: Sao Tome	Peter Andolfatto lab

Table S2. Drosophila species and populations used in study. Clade, population of origin, number of samples, and reference are given in each column.

Species	Clade	Population	# Samples used	Source
D. melanogaster	*melanogaster*	North America (RAL)	100	MacKay et al. 2012; Lack et al 2015
D. melanogaster	*melanogaster*	Africa (ZI)	100	Pool et al 2012; Lack et al 2015
D. simulans	*simulans*	North America (CA)	170	Signor 2018
D. sechellia	*simulans*	Africa (Praslin)	23	Schrider et al. 2018
D. mauritiana	*simulans*	Africa (Mauritius)	15	Garrigan et al. 2014 Serrato-Capuchina et al. 2021 Meany et al. 2019
D. santomea	*yakuba*	Africa (São Tomé)	7	Turissini and Matute 2017; Bachtrog et al 2007;
D. teissieri	*yakuba*	Africa (Bioko, Zimbabwe, Gabon)	11	Turissini and Matute 2017

Table S3. Minimum and mean π/bp computed in non-overlapping 20Kb windows on the X chromosome across species. The fourth column shows the minimum π/bp on the X as a percent of the average π/bp on the X chromosome. In *D. simulans* the window with lowest diversity is 0.5% of the average π/bp while on *D. santomea* it is 10% of the average π/bp.

Species	min(π/bp)	mean(π/bp)	% of mean(π/bp)
D.mauritiana	0.00004	0.006	0.7%
D. melanogaster (RA)	0.0002	0.003	6.7%
D. melanogaster (ZI)	0.0006	0.008	7.5%
D.santomea	0.0005	0.005	10.0%
D. sechellia	0.000002	0.0003	0.7%
D. simulans	0.00002	0.004	0.5%
D. teissieri	0.0003	0.01	3.0%

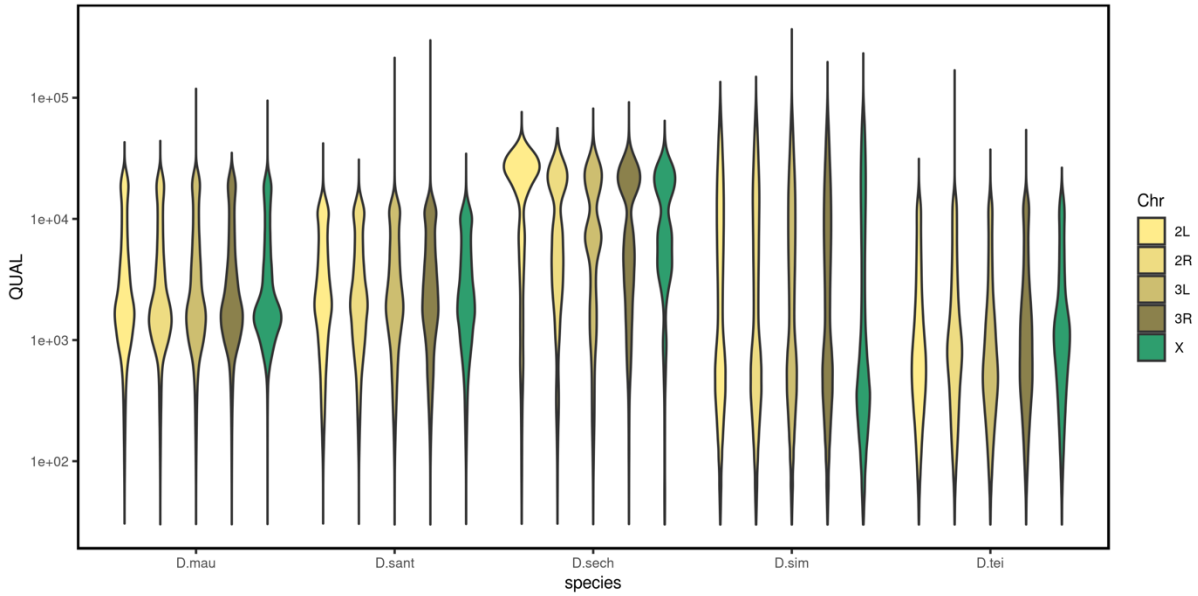


Figure S1. Phred-scaled quality score across species. The distribution is similar across autosomes and the X chromosome for *D. mauritiana, D. santomea. D. sechellia, D. simulans* and *D. teissieri*. The y-axis is log-scaled for better visualization.

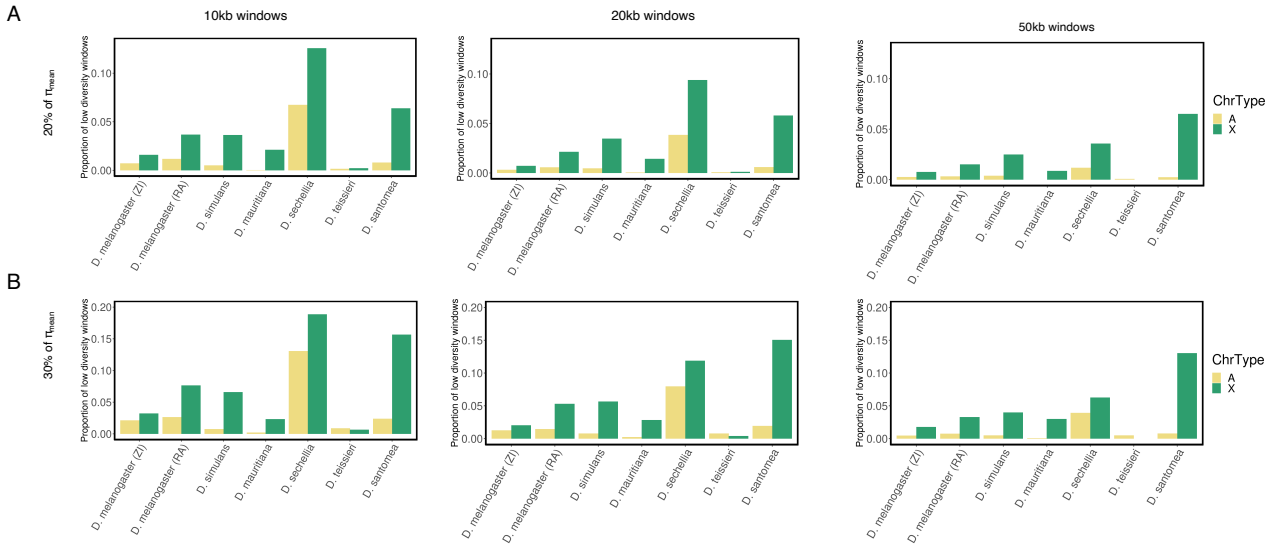


Figure S2. Low diversity windows on the X versus autosomes as a function of the low diversity thresholds and window size. Proportion of low diversity windows on X versus autosomes (top row, green and yellow bars). Three different analysis windows were tested: 10kb, 20kb and 30kb windows, as well as two different diversity thresholds: **(A)** 20% and **(B)** 30% of the chromosome's π/bp average. We consistently see a higher proportion of low diversity windows on the X compared to autosomes. This figure is analogous to **Fig. 2B** in the main text.

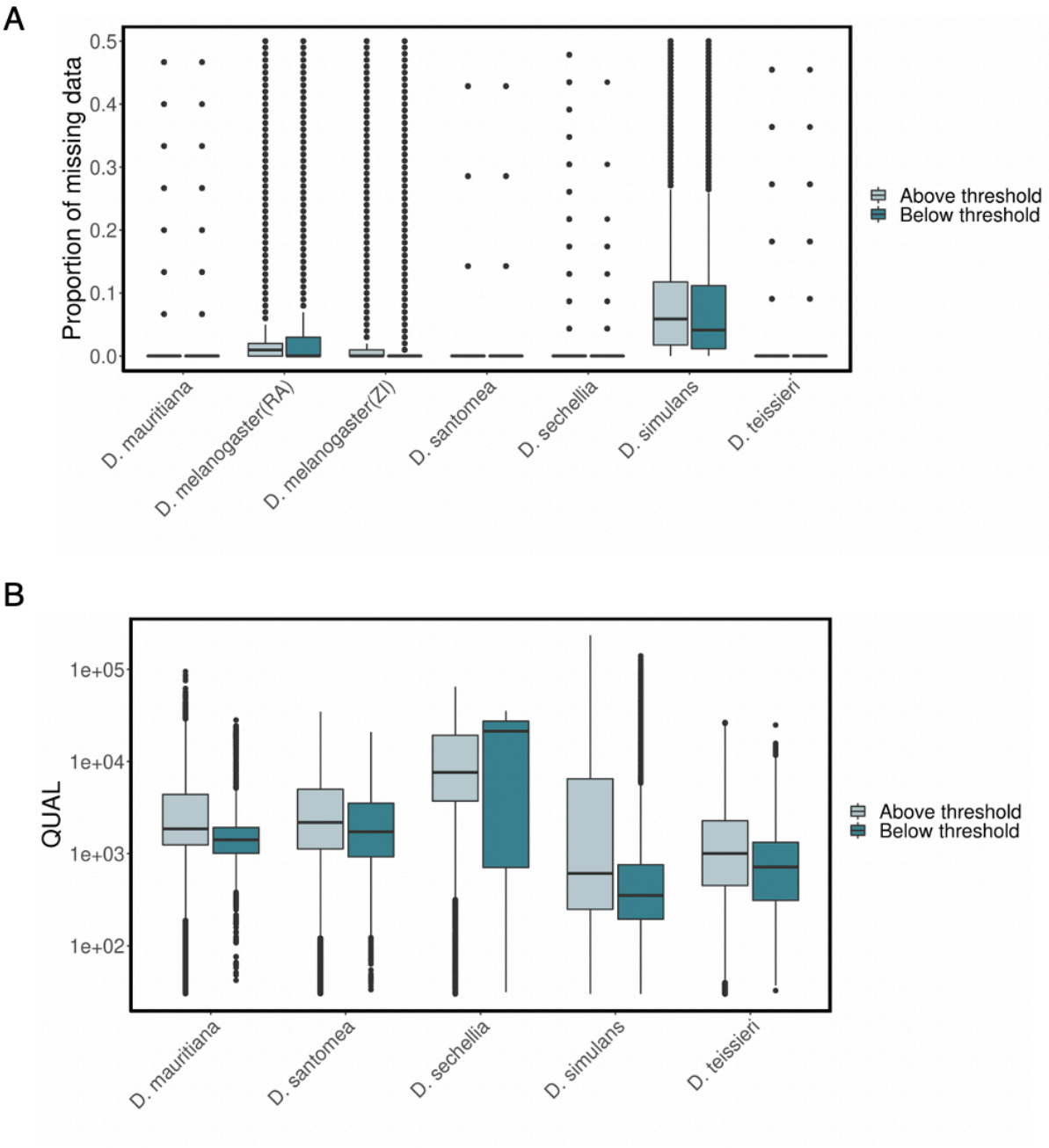


Figure S3. Proportion of missing data (A) and Phred-scaled quality score (B) in windows above and below the low diversity threshold for the X chromosome. The y-axis in (B) is log-scaled for better visualization.

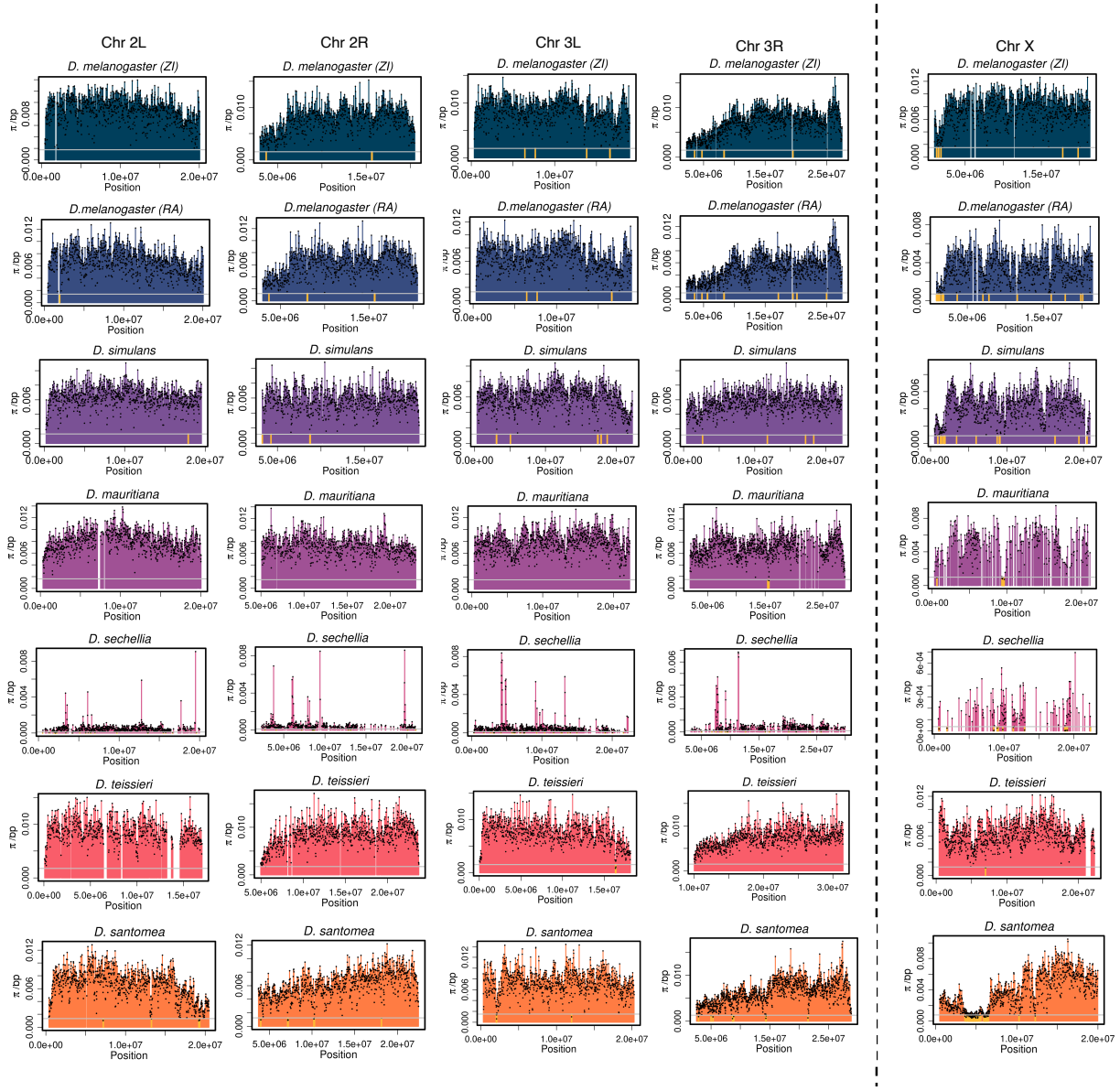


Figure S4. Genome-wide diversity in six species. π/bp in 20kb bins across four autosomal arms and the X chromosome of six *Drosophila* species and two *D. melanogaster* populations from Raleigh (RA) and Zambia (ZI). Windows where diversity is below 20% of the chromosomal average (grey horizontal line) are highlighted in gold.

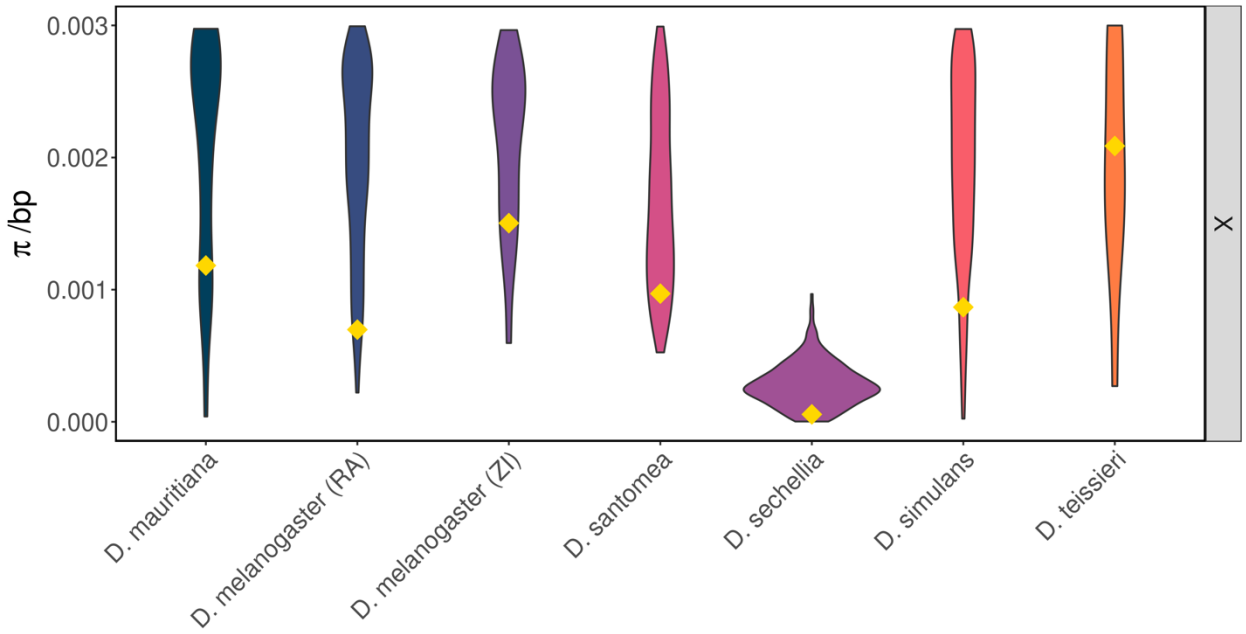


Figure S5. Distribution of π/bp for the X chromosome across species. Violin plots for π/bp for each species' X chromosome are plotted with y-axis going up to π/bp=0.03 for better visibility of the lower tail of the distribution. The low diversity threshold is shown as the yellow diamond on the tail of each violin plot and represents 20% of the mean value of the respective distribution. The lower tail of each distribution extends well below the low diversity threshold for all species.

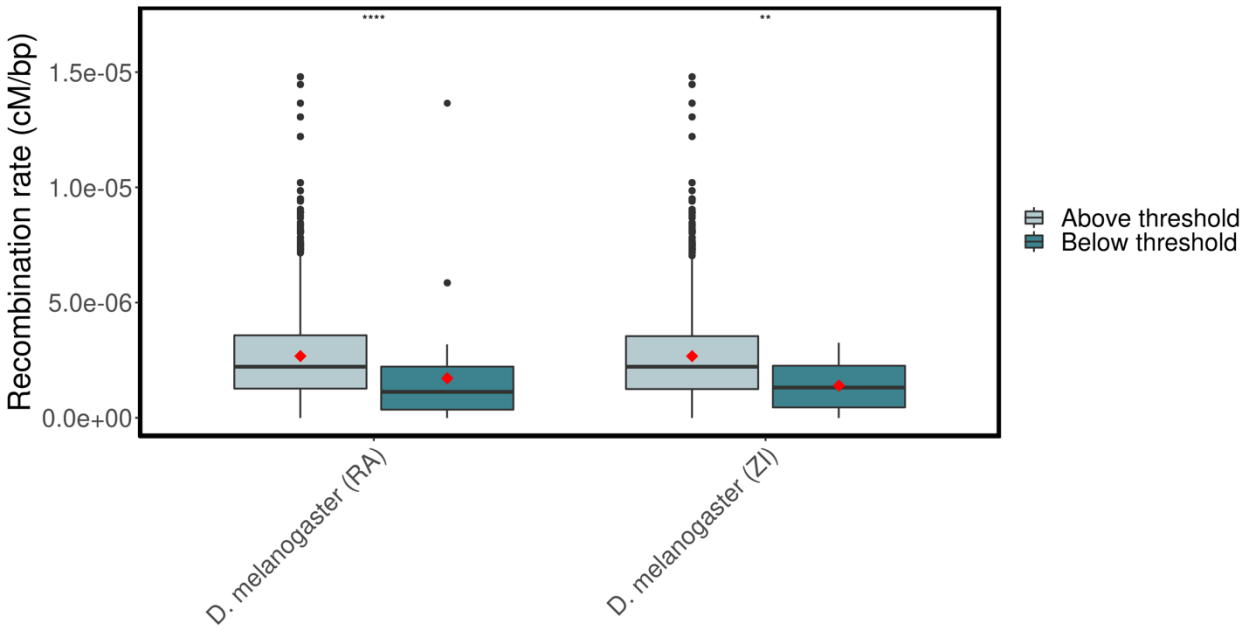


Figure S6. *D. melanogaster* recombination rates (cM/bp) in regions below and above the low diversity threshold. Recombination rates for *D. melanogaster* were obtained from the Comeron et al 2012 recombination map. The red data points correspond to the average recombination rate in each category (r = 1.71 x 10-6 cM/bp (RA),1.40 x 10-6 cM/bp (ZI) in low diversity windows and r = 2.68 x 10-6 cM/bp (RA), 2.67 x 10-6 cM/bp (ZI) in high diversity windows). We find a lower recombination in low diversity windows (dark blue) compared to windows above (light blue) the low diversity threshold. The asterisks represent a significant difference in recombination rate using a Wilcoxon rank-sum test.

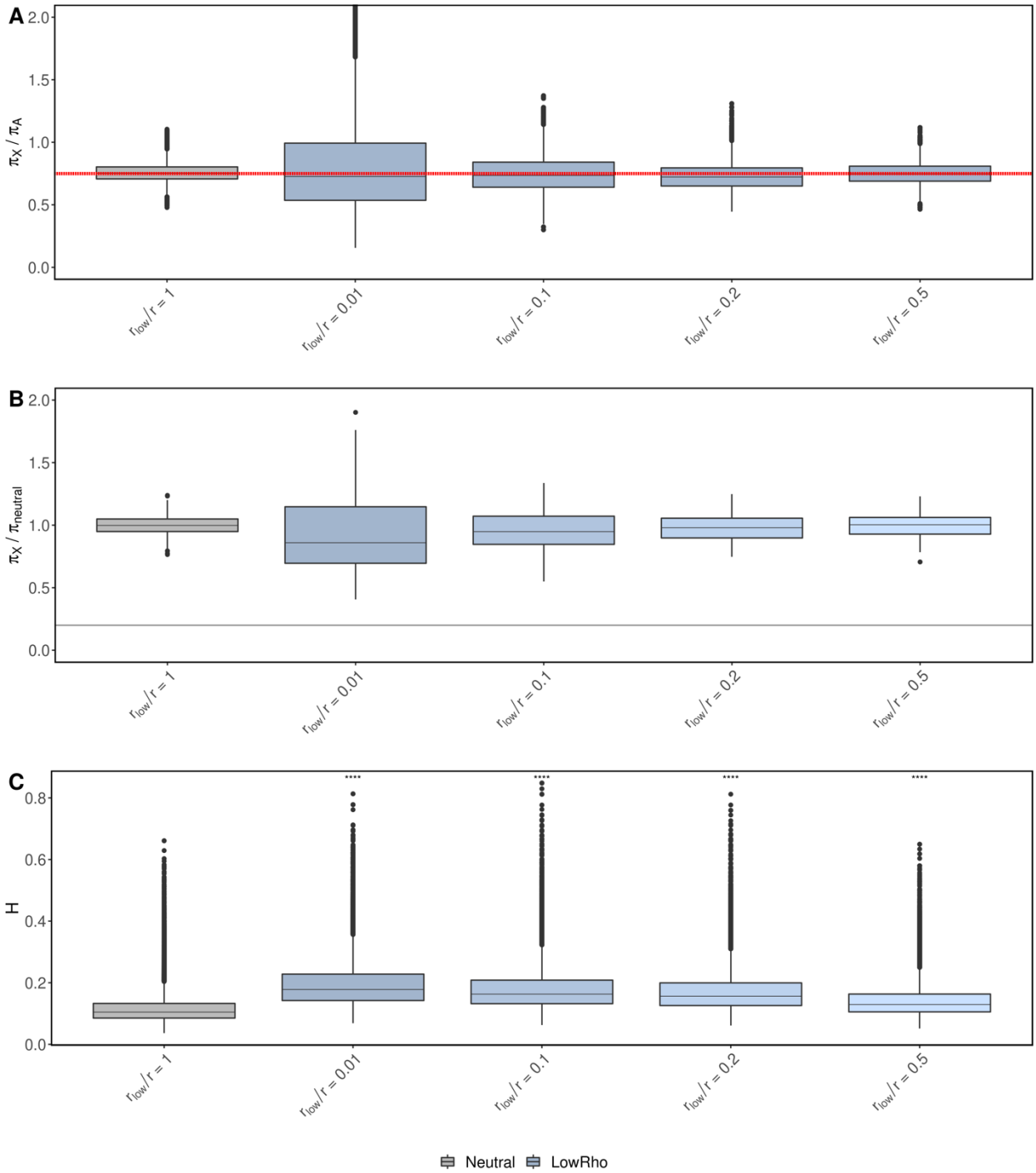


Figure S7. Windows of low recombination can elevate haplotype homozygosity but cannot reduce diversity below 20% of $\pi_{neutral}$. Models of regions with low recombination rate (r_{low}) such that $r_{low}/r=0.01,0.1,0.2$ and 0.5, where $r=5e\text{-}9cM/bp$. These regions of low recombination cannot reduce X-linked diversity below the π_X/π_A =0.75 expectation (**A**) or below the low diversity threshold **(B)** but can elevate haplotype homozygosity compared to neutrality **(C).**

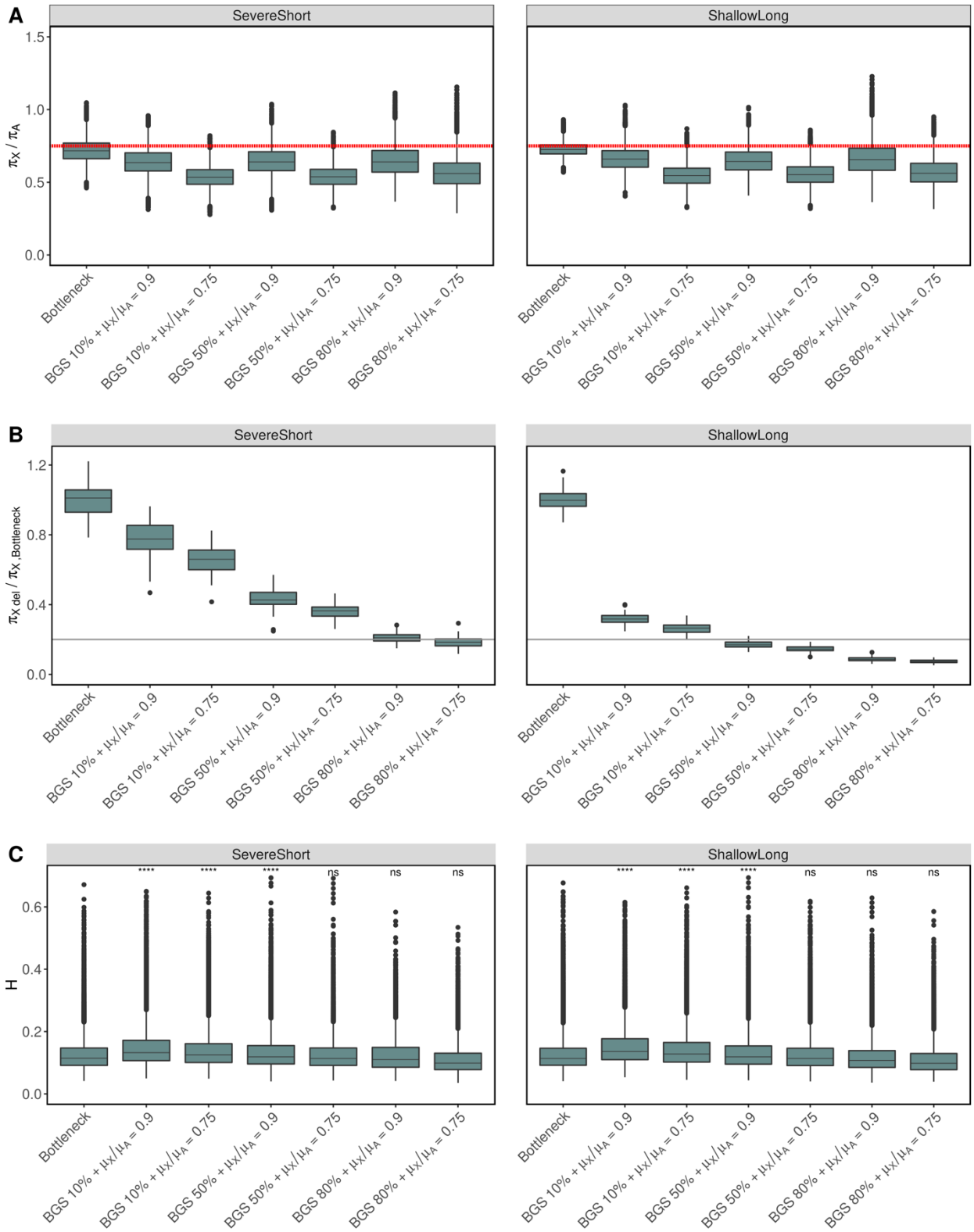


Figure S8. Background selection combined with bottlenecks and low X-linked mutation rate for *r*= 2.5e-7cM/bp. The models considered models considered include three BGS models each

combined with a lower X linked mutation rate (μ_X/μ_A =0.9, μ_X/μ_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 -0.000133 and 0.35, respectively (Huber *et al.* 2017). The left column corresponds to a severe short bottleneck model and the left to a shallow long bottleneck model (**Methods**). For each model, we computed: **(A)** π_X/π_A across all models, where the red dashed line corresponds to the expected π_X/π_A = 0.75 value in the baseline 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. The mean $\pi_{X,Bottleneck}$ represents the average diversity in the chromosome. The solid gray line is 20% of $\pi_{X,Bottleneck}$. **(C)** Haplotype homozygosity across all models and results from a one-sided Wilcoxon rank-sum test for elevation of H in the bottleneck+BGS+Low μ_X scenario compared to a neutral bottleneck model.

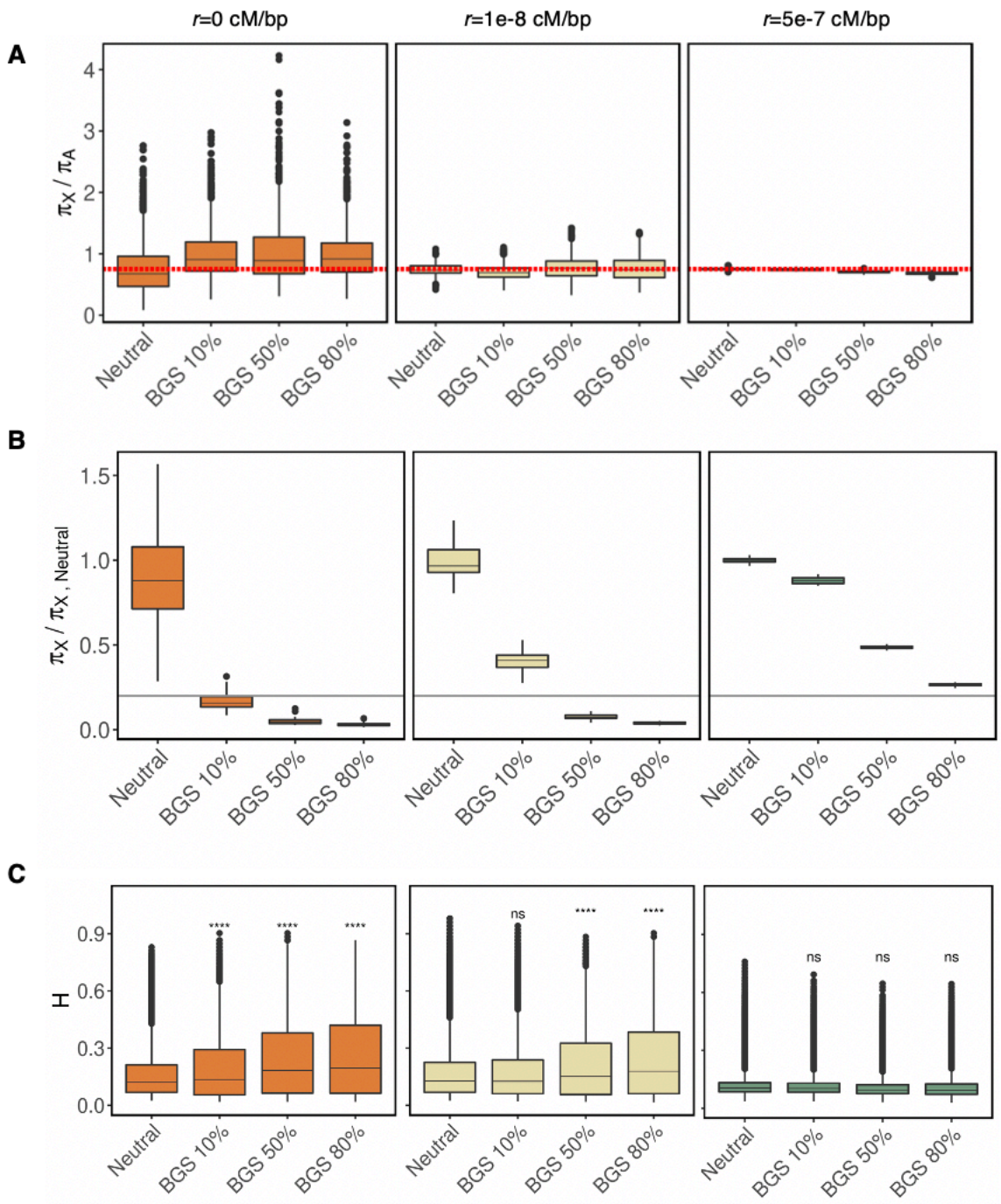


Figure S9. Nucleotide diversity (π) and haplotype homozygosity (H) in 1Mb simulated chromosomes with BGS, varying recombination rate (*r*). The models considered include a neutral model with a constant $N_e=10^6$ and no sex bias, and, three BGS models where we varied the proportion of deleterious mutations (10, 50 and 80%). The DFE for the deleterious selection coefficient (s_d) was gamma distributed with mean and shape parameter -0.000133 and 0.35, respectively. Simulated three different recombination rates: *r*=0,1e-8 and 5e-7 cM/bp. 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 π_X is the average π in the scenario being considered on the X-axis, and $\pi_{X,Neutral}$ is the average π in the baseline neutral model. 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 one-sided Wilcoxon rank-sum test.

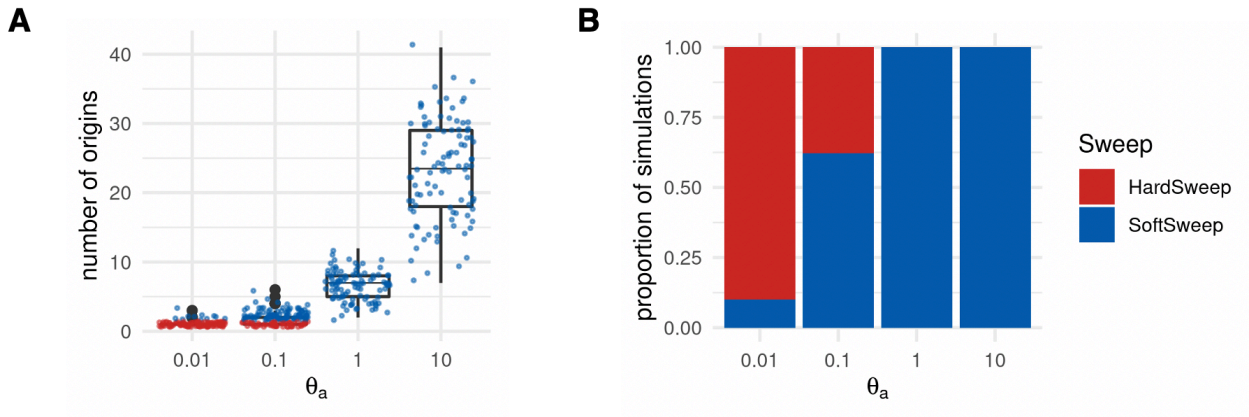


Figure S10. Number of origins and proportion of hard/soft sweep simulations as a function of θ_a. Soft sweeps (blue) were defined as simulations in which the sample had >1 mutational origin for the adaptive mutation. Hard sweeps (red) were defined as simulations with one mutational origin for the adaptive mutation. **(A)** Distribution of the number of origins for the adaptive mutation as a function of θ_a. **(B)** Proportion of hard versus soft sweeps in simulations as a function of θ_a.

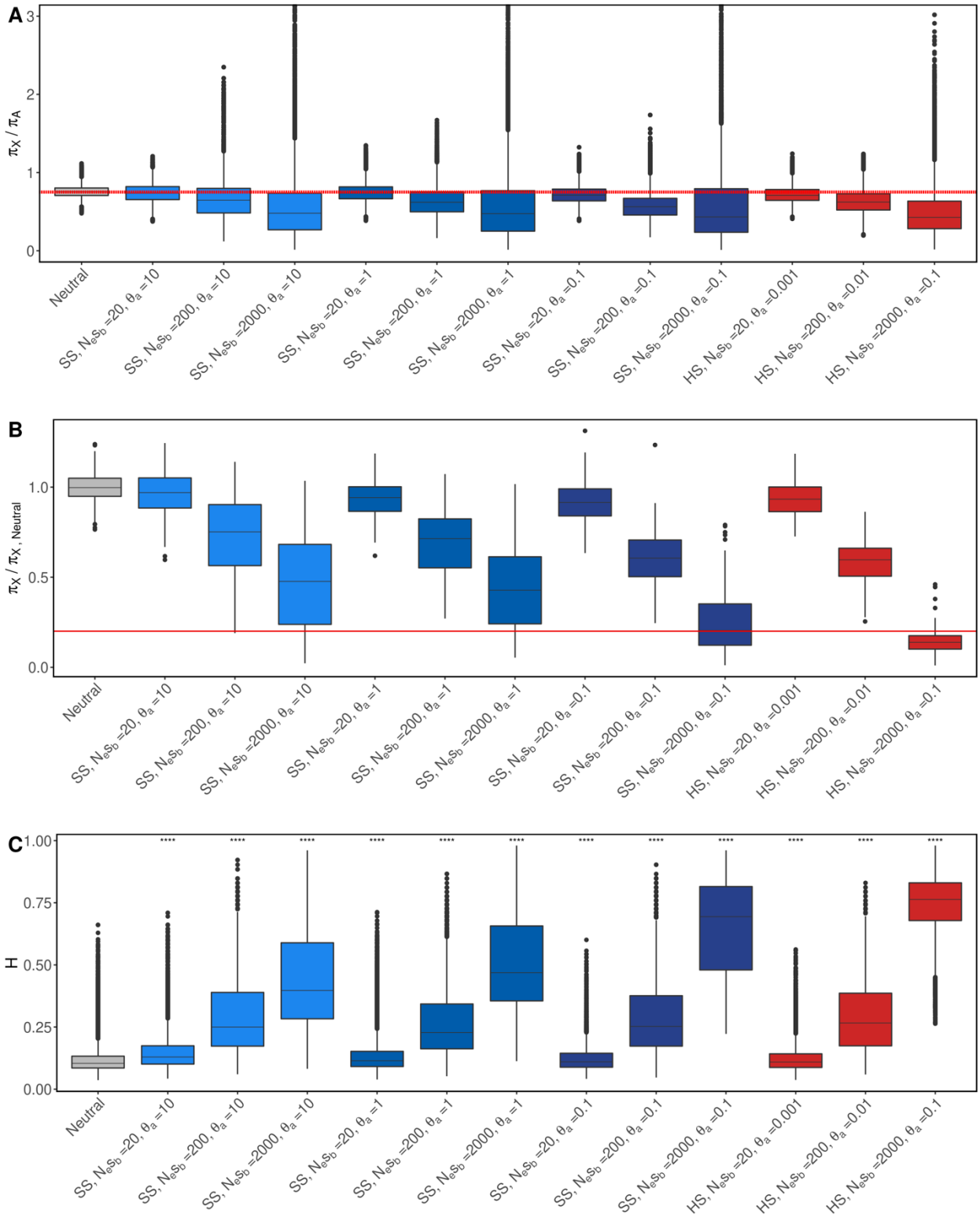


Figure S11. Diversity (π) and haplotype homozygosity (H) in hard and soft sweeps. Models of hard sweeps (red; θ_a =0.01) and soft sweeps varying softness (blue; θ_a =0.1,1 and 10) with three

different selection strengths ($N_e s_b$=20, 200, 2000). 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)** $\pi_X/\pi_{neutral}$ where $\pi_{neutral}$ is the average π in the baseline neutral model and the solid gray line is 20% of $\pi_{neutral}$; and **(C)** haplotype homozygosity. The asterisks represent a significant elevation in haplotype homozygosity relative to complete neutrality using a one-sided Wilcoxon rank-sum test. Both hard and soft sweeps can decrease π_X/π_A <0.75 and elevate haplotype homozygosity. When selection is strong ($N_e s_b$=2000) it is possible for both hard and soft sweeps to reduce π_X below 20% of $\pi_{neutral}$. However, the proportion of hard sweeps simulations below this threshold is higher (0.82) compared to soft sweeps (0.48 for θ_a=0.1, 0.19 for θ_a=1 and 0.18 for θ_a=10), making hard sweeps more likely.

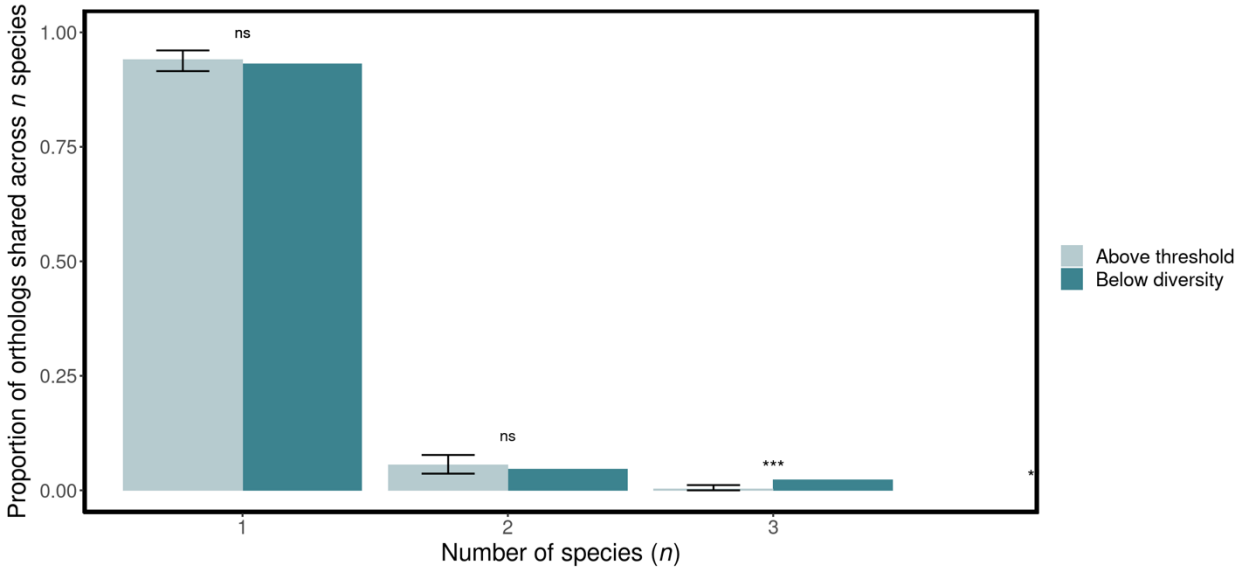


Figure S12. Proportion of orthologs shared across species in windows below and above the low diversity threshold. We obtained 100 random samples of orthologs found in windows above the low diversity threshold and obtained the mean number of orthologs shared across *n=1,2 or 3* species (light blue) as well as the corresponding 95% confident intervals for each case. The case of *n=1* represents orthologs found in a single species only. For *n=3,* we found that the proportion of orthologs shared across species is higher in low diversity windows. Comparisons between the two *D. melanogaster* populations were excluded from the analysis in order to emphasize between-species dynamics.

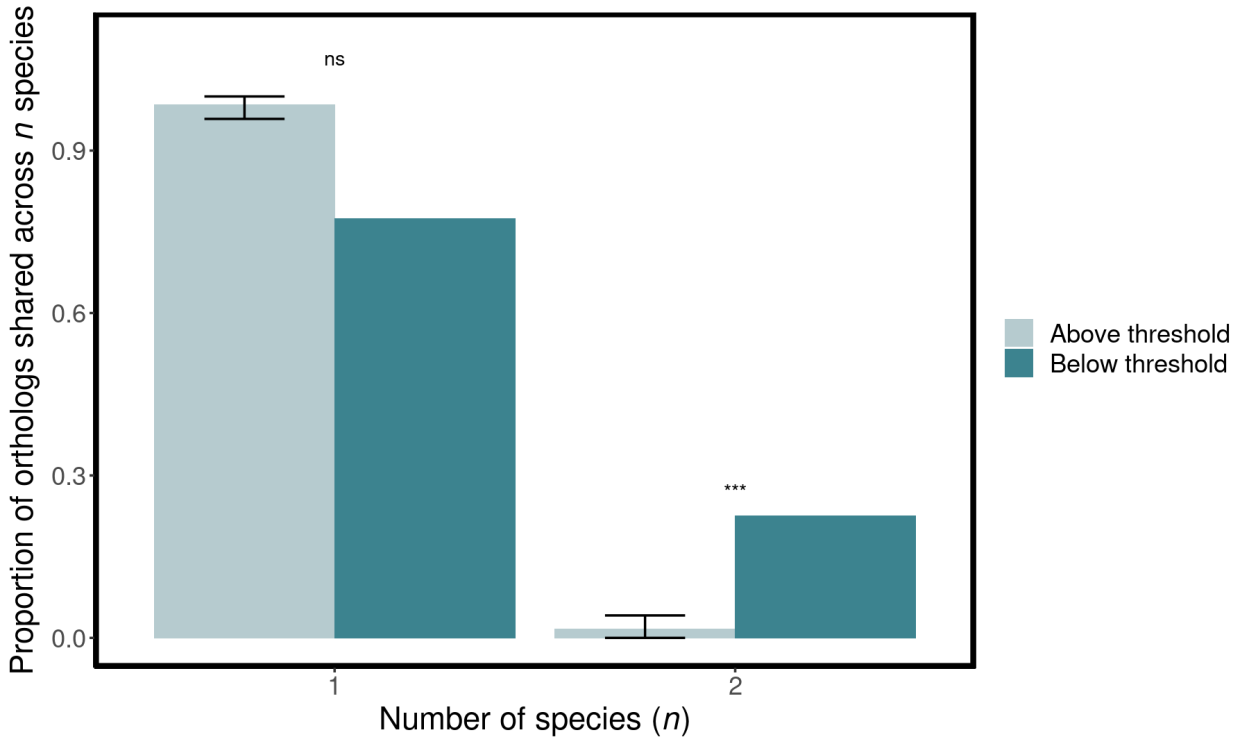


Figure S13. Proportion of orthologs shared across species in windows below and above the low diversity threshold in *D. melanogaster* (ZI) and *D. melanogaster* (RA) only. We obtained 100 random samples of orthologs found in windows above the low diversity threshold and obtained the mean number of orthologs shared across *D. melanogaster* populations (light blue) as well as the corresponding 95% confident intervals. We show the comparison between *D. melanogaster* populations in order to emphasize processes that may be influencing different populations of the same species.

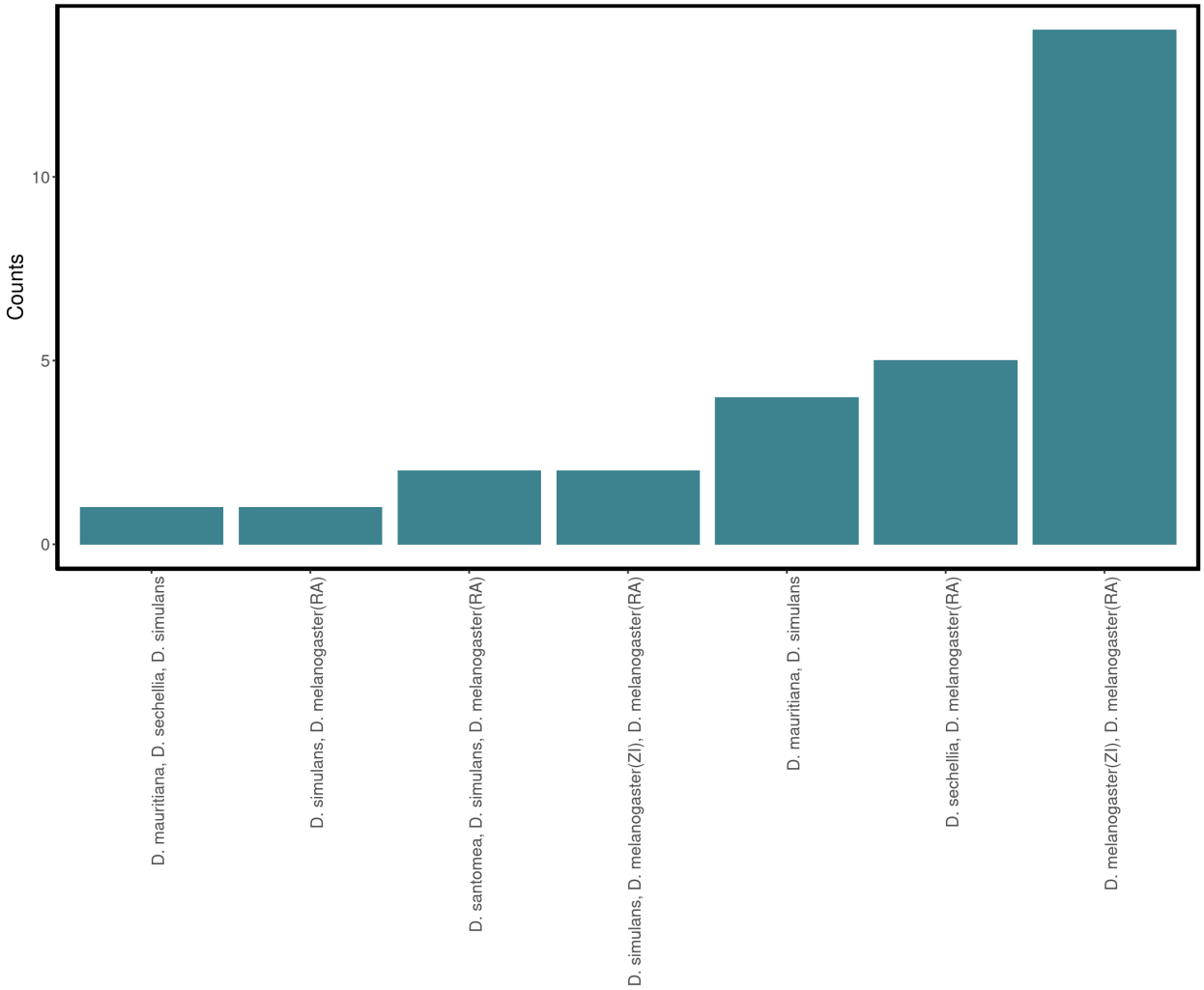


Fig S14. Number of orthologs shared across species. The x-axis shows the clusters of species that have orthologs in common. The *D. melanogaster* populations have the highest number of overlapping orthologs. Overlap between three or more species is less common.

Text S1: Overlapping orthologous gene groups on the X chromosome across species

To investigate whether the low diversity regions observed across species on the X chromosome correspond to similar functions under selection across species, we identified orthologous gene groups present in two or more species for regions both below and above the low diversity threshold (**Methods**). We found that, excluding *D. melanogaster* (ZI) versus *D. melanogaster* (RA) comparisons, the proportion of orthologous genes that are present across three species was significantly higher in low diversity regions on the X compared to the rest of the chromosome (**Fig. S12**). We did not find a significant difference for orthologous genes found in two species. Specifically, we observed that 5% and 2% of orthologs in low diversity windows were shared across 2 or 3, respectively. In contrast, for regions above the threshold, 6% and 0.3% of orthologs where identical across 2 or 3 species, respectively. Moreover, when we included all populations, we found the highest overlap was between *D. melanogaster* (ZI) and *D. melanogaster* (RA) (14/ 43 orthologous gene groups found in at least two species; see **Fig. S13, Fig. S14**). This finding aligns with our expectations since *D. melanogaster* (ZI) and *D. melanogaster* (RA) are two populations of the same species, suggesting that the selection process may have originated in the ancestral population and persisted in the derived population. Despite the several instances of shared targets of selection across species, the majority of genes found in low-diversity regions were found in a single species (**Fig S12, Fig. S13**), indicating that distinct genes are targeted by selection independently.