# Hill-Robertson Interference Causes Reduced Genetic Diversity on a Young Plant Y-chromosome Following Recent Recombination Suppression

**Josh Hough**[*,†,1], **Wei Wang**[†], **Spencer C.H. Barrett**[†] **and Stephen I. Wright**[†]

[*]Department of Plant Sciences, University of California, Davis, [†]Department of Ecology and Evolutionary Biology, University of Toronto

**ABSTRACT** X and Y chromosomes differ in effective population size ($N_e$), rates of recombination, and exposure to natural selection, all of which can affect patterns of genetic diversity. On Y chromosomes with suppressed recombination, selection is expected to eliminate neutral variation and reduce the $N_e$ of Y compared to X chromosomes or autosomes. However, non-selective factors including female biased sex ratios and high variance in male reproductive success can also reduce Y-linked $N_e$, making it difficult to infer the causes of low Y-diversity from natural populations. Here, we investigate the factors affecting levels of polymorphism during sex chromosome evolution in *Rumex hastatulus* (Polygonaceae), a dioecious plant with young sex chromosomes. Strikingly, we find that neutral diversity for genes on the Y is on average 2.1% of the value for their homologues on the X, corresponding to a chromosome-wide reduction of 93% compared to the neutral expectation. We demonstrate that the magnitude of this diversity loss is inconsistent with a reduced male $N_e$ caused by neutral processes including female-biased sex ratios and high variance in male reproductive success. Instead, using forward simulations and estimates of the distribution of fitness effects of deleterious mutations, we show that diversity loss on the Y can be explained by purifying selection acting over a large number ($\geq$ 800 Kb) of genetically-linked sites. Our results are in agreement with theory on "interference selection", and provide empirical evidence that this effect can substantially reduce neutral diversity on a non-recombining chromosome. Given the recent origin of *R. hastatulus* sex chromosomes (15MYA), our results imply that Y-chromosome degeneration in the early stages may be largely driven by such interference effects rather than by positive selection for gene silencing followed by neutral genetic drift.

**KEYWORDS** Sex Chromosome Evolution; Nucleotide Diversity; Recombination; Deleterious Mutations

## Introduction

Morphologically distinct sex chromosomes have evolved multiple times independently in both plants and animals (Westergaard 1958; Ohno 1967; Bull 1983; Charlesworth 1991, 2015). Despite clear biological differences between these kingdoms, X and Y chromosomes in both lineages have undergone similar genetic changes. For example, in both groups the loss of recombination between X and Y chromosomes has been associated with an accumulation of deleterious mutations and a grad-ual loss of genes from the Y chromosome (Hough *et al.* 2014; Bergero *et al.* 2015; Bachtrog 2013), and in some species, the degeneration of the Y has led to the evolution dosage compensation of the X chromosome (Charlesworth 1996; Muyle *et al.* 2012; Mank 2013; Papadopulos *et al.* 2015). The independent evolution of these phenomena in such taxonomically distant groups suggest that general evolutionary mechanisms may be involved, but inferring the causes of molecular evolution and patterns of polymorphism in genomic regions that lack recombination is a longstanding challenge for both theoreticians and experimentalists (Charlesworth 1978; Feldman *et al.* 1980; Barton 1995; Charlesworth 1996; Otto and Feldman 1997; Charlesworth and Charlesworth 2000; McVean and Charlesworth 2000a).

One fundamental difference between the X and Y chromosomes is that there are 1/3 as many Y-linked gene copies as

X-linked ones in a diploid population. Thus, genes on the Y chromosome are expected to experience an effective population size ($N_e$) that is 1/4 that of autosomal genes, whereas the $N_e$ for genes on the X chromosome should be 3/4 that of autosomal genes (assuming an equal number of reproducing females and males). The lowered $N_e$ of the Y chromosome implies that that the equilibrium level of neutral polymorphism - proportional to the product of $N_e$ and the neutral mutation rate, $\mu$ - should be lower for Y-linked genes than for their X-linked counterparts. In the absence of recombination, genes on the Y chromosome are also expected to be in strong linkage disequilibrium, making them vulnerable to diversity loss due to selection against strongly deleterious mutations (background selection) and selective sweeps of strongly beneficial mutations (genetic hitchhiking). Furthermore, the build-up of linkage disequilibrium between selected mutations on the Y means that selection will act non-independently across the chromosome, such that selection at a focal site may "interfere" with selection at sites with which it is linked (Hill and Robertson 1966). A large body of work has now shown that in theory such "selective interference" effects can substantially reduce both the efficacy of selection and the level of neutral variability (Fisher 1930; Muller 1964; Hill and Robertson 1966; McVean and Charlesworth 2000b; Kaiser and Charlesworth 2009; Good et al. 2014). These arguments all suggest that non-recombining Y chromosomes should harbor a lower amount of neutral genetic variability than predicted based on the number of Y chromosomes in a population, but the extent to which background selection, genetic hitchhiking, or selective interference have affected chromosome-wide levels of diversity in natural populations, and the relative importance of these processes, is not well-understood.

In addition to reduced diversity arising from selection, in species with female-biased sex ratios or extensive male-male competition, high variance in male reproductive success is also expected reduce the $N_e$ experienced by genes on the Y chromosome (Caballero 1995; Charlesworth 2001; Laporte and Charlesworth 2002; Pool and Nielsen 2007; Ellegren 2009), suggesting that inferences about the effects positive of purifying selection need to be distinguished from these neutral processes. Because variance in male reproductive success differentially affects the $N_e$ of X, Y, and autosomal chromosomes (Kimura and Crow 1964; Nomura 2002), evidence for this can therefore be obtained by comparing levels of neutral diversity on X and Y, relative to autosomes. For example, high variance in male reproductive success causes a reduction in the Y/A diversity ratio, but an increase in the X/A ratio. Based on such comparisons, studies in humans, for example, have suggested that the inflated X/A ratio is due to a historical excess of breeding females compared to males (Hammer et al. 2008) (and see (Bustamante and Ramachandran 2009; Hammer et al. 2010; Cotter et al. 2016)).

Despite widespread interest in determining the evolutionary factors affecting neutral diversity on sex chromosomes (Ellegren 2011; Bachtrog 2013), we know very little about the influence of sex ratio variation or linked selection in determining levels of diversity on more recently evolved sex chromosomes. The time scales over which these different effects are likely to be important is therefore not well understood. In humans, estimates of Y-linked diversity are considerably lower than predicted under neutral models, and simulations suggest that levels of diversity are consistent with strong purifying se-

lection (Wilson Sayres et al. 2014). However, given that human sex chromosomes evolved from autosomes ~200 million years ago (MYA) (Lahn and Page 1999a; Ross et al. 2005), is not immediately obvious whether purifying selection might have such strong effects on Y chromosomes that evolved *do novo* from autosomes over much more recent evolutionary time (e.g., within the last ~20 MYA in the case of plants (Charlesworth 2015)). Y-linked diversity loss might on the one hand be expected to be lower in younger if there has been widespread gene silencing, as discovered on the *Drosophil miranda* neo-Y chromosome (Zhou and Bachtrog 2012). On the other hand, simulations of strong selection models (background selection and genetic hitchhiking) suggest that these processes may have the greatest effects during the earliest stages of sex chromosome evolution, before the Y has lost many of its genes (Bachtrog 2008). Moreover, there is clear evidence from simulation studies that even weak purifying selection, if acting non-independently across large number of sites, can generate substantial deviations from neutrality, whereas classic background selection theory breaks down in such cases (McVean and Charlesworth 2000b; Comeron and Kreitman 2002; Kaiser and Charlesworth 2009; Good et al. 2014). Given that a large number of selected sites are likely to be in linkage disequilibrium on a recently evolved Y chromosome, such "interference selection" *sensu* (Good et al. 2014) is a good candidate model for exploring the evolutionary dynamics of young plant sex chromosomes.

To investigate the factors affecting nucleotide diversity in the early stages of sex chromosome evolution, we analyzed neutral polymorphism levels on X, Y, and autosomal chromosomes in the plant *R. hastatulus* (Polygonaceae). This species is a dioecious annual with heteromorphic X and Y chromosomes that evolved ~15 MYA (Quesada del Bosque et al. 2011; Grabowska-Joachimiak et al. 2015; Navajas-Pérez et al. 2005), making Y chromosomes in this species over 100 million years younger than the highly degenerated Y chromosomes in mammals (Lahn and Page 1999b; Ross et al. 2005). *Rumex hastatulus* has also received particular attention because of the occurrence of an interesting polymorphic in sex chromosome system, in which both XY and $XY_1XY_2$ males occur in geographically distinct populations ( so-called "chromosomal races") (Smith 1963). The $XY_1XY_2$ sex chromosome system in this species is thought to have originated through an X-autosome fusion, with the XY system maintaining the ancestral chromosome complement (Smith 1964). Despite the recent origin of sex chromosomes in both races, there is evidence that both Y's have undergone gene loss and functional deterioration (Hough et al. 2014). Here, to simplify our comparison of polymorphism levels on X, Y, and autosomes, we focus only on the XY system with the ancestral chromosome configuration, as it is more likely to be at equilibrium than the neo-Y chromosome system that experienced the fusion event.

Of particular importance for our study, *R. hastatulus* populations have also been found to consistently exhibit female-biased reproductive sex ratios, with a mean sex ratio of $N_f/(N_m + N_f) = 0.6$ (Pickup and Barrett 2013). Female-biased sex ratios are not uncommon in dioecious plants with heteromorphic X and Y chromosomes (Field et al. 2013; Hough et al. 2013), and their occurrence in *R. hastatulus* provides an excellent opportunity to investigate both the demographic and selective factors contributing to sex-linked variability.

## Materials and Methods

### Population Samples and Sex-Linked Genes

We analyzed sex-linked and autosomal genes identified from Illumina RNA sequence data from 12 males and 12 females (1 male and 1 female from each of 6 populations). Samples were collected in 2010 from throughout the native range of *R. hastatulus* (locations in Table S1), and plants were grown in the glasshouse at the University of Toronto from seeds collected from open-pollinated females. We extracted RNA from leaf tissue using Spectrum Plant Total RNA kits (Sigma-Aldrich). Isolation of mRNA and cDNA synthesis was conducted according to standard Illumina RNAseq procedures, and sequencing was conducted on two Illumina HiSeq lanes with 150-bp end reads at the Genome Quebec Innovation Center. Reads from these samples were mapped to the *R. hastatulus* reference transcriptome (Hough *et al.* 2014) using the BurrowsWheeler Aligner (Li and Durbin 2010), followed by Stampy (Lunter and Goodson 2011). We used Picard tools (http://picard.sourceforge.net) to process mapping alignments for the Genome Analysis Toolkit (McKenna *et al.* 2010) variant calling software, and subsequently removed genes with low coverage (<10x) and low Phred Quality Scores <20. The population samples analyzed here were previously reported in Hough et al. (2014), where they were used to validate the ascertainment of sex-linked genes identified through segregation analysis, and raw sequences are available from the GenBank Short Read Archive under accession no. SRP041588. Here, to consider sex linked genes that were identified in all of our sequenced population samples, we focused on the previously described set of 460 X/Y genes for which a Y homolog was found in both the Texas and North Carolina races (i.e., X/Y genes where the Y copy was inferred to be on the $Y_1$ chromosome).

### Autosomal Genes

In evaluating evidence for nucleotide diversity differences between X and Y chromosomes, it is important to distinguish between reduced Y-linked diversity, and the possibility that X-linked diversity is elevated above the level predicted from a neutral model. To do this, we normalized our sex-linked diversity estimates by autosomal diversity, and compared empirical X/A and Y/A nucleotide diversity ratios to those predicted from neutral models and from simulations (described below). Because the criteria for ascertaining autosomal loci in Hough et al. (2014) were based on identifying four segregating SNPs per locus, and since this set of genes is likely to be higher in diversity than the average autosomal gene, here we instead used the larger set of all non-sex linked (putatively autosomal) genes as our autosomal reference. We filtered genes in this set to remove any genes that may have been sex-linked but were not identified as such by Hough et al.'s conservative ascertainment criteria. In particular, we removed: (i) any genes in which there was evidence for at least one SNP with a sex-linked segregation pattern, (ii) any genes where SNPs showed fixed heterozygosity in males and fixed homozygosity in females, (iii) genes with less than 10X coverage or greater than 100X coverage from independently obtained genomic coverage data (to filter out duplicates or genes with highly repetitive sequences), and (iv) any genes containing SNPs with large (>0.4) allele frequency differences between males and females. Finally, we removed genes with fewer than 100 synonymous sites to avoid biasing our results toward genes that may have been particularly short due to assembly problems. This filtering resulted in a final set of 12,356 autosomal genes.

### Phasing X and Y alleles

To estimate polymorphism for X and Y sequences separately in males, it is necessary to infer the phase of SNPs in sex-linked transcripts in males. In previous work, phasing alleles on *R. hastatulus* sex chromosomes was achieved using segregation analysis from a genetic cross. Here, to phase SNPs from population samples where such segregation data was unavailable, we used HAPCUT (Bansal and Bafna 2008), a maximum-cut based algorithm that reconstructs haplotypes using sequenced fragments (Illumina read data) from the two homologous chromosomes to output a list of phased haplotype blocks containing the SNP variants on each chromosome. Because the resulting haplotype blocks produced by HAPCUT contained SNPs that were phased relative to each other, but not designated to either the X or Y chromosome, we assigned individual variants to X or Y by independently identifying fixed X-Y differences within each haplotype block (i.e., sites where all females were homozygous, and all males were heterozygous). Identifying such fixed differences within phased haplotype blocks enabled us to then infer the correct phase (X or Y) of the polymorphisms from HAPCUTs output. In particular, this was done by matching the phase of fixed X-Y differences with their neighboring polymorphic sites: when a fixed X-Y difference occurred in the same phased haplotype block as a polymorphic site, then the polymorphic variants in that block were assigned to either X or Y based on the known phase of the fixed difference with which they were matched. SNPs that were identified outside of phased blocks, or in blocks without fixed X-Y differences, were recorded as missing data. Finally, we filtered out SNPs with coverage < 60, and those within a distance of 10bp or less from indels. This filtering procedure resulted in alignments of X and Y sequences for 372 sex-linked genes.

We further validated the results of HAPCUTs allele phasing by comparing the accuracy of this method with the phasing-by-segregation method that was conducted in Hough et al. (2014). To do this, we first phased the sequence data from parents and their progeny using HAPCUTs algorithm (using the same parameters as for the population data), and then identified cases where SNPs were inferred on the Y chromosome by HAPCUT, but where the true level polymorphism, obtained from the genetic cross, was zero. We identified 7% of sex-linked genes that either had phasing errors genotyping errors. This corresponds to a SNP error rate estimate of 1.7 x 10-4. Note that this rate is very low relative to population-based estimates of polymorphism on the X and autosomes (Table 1), and therefore should have minimal effects on our estimation of the X/A ratio. However, because this rate is high relative to the expected level of polymorphism on the Y chromosome, we conducted a further filtering step for Y-linked SNPs and identified false positive SNP calls arising from: (i) phasing errors caused by gene duplicates (more than two haplotypes), (ii) polymorphisms around indels, and (iii) genotyping errors caused by low Y-expression. This final filtering was conducted by manually checking each individual putative polymorphism on the Y chromosome using IGV (Robinson *et al.* 2011).

### Estimating nucleotide diversity on sex chromosomes and autosomes

For each locus in our analysis, we calculated Watterson's (1975) estimator of the population parameter $\theta = 4N_e\mu$, where $N_e$ is

the effective population size, and $\mu$ is the mutation rate (Watterson 1975), using a modified version of the Perl program Polymorphurama (Bachtrog and Andolfatto 2006). To compare sex-linked and autosomal loci, we calculated the average value of $\theta$, weighted by the number of synonymous sites in each gene (Figure 2). We obtained 95% confidence intervals for X/A and Y/A ratios by bootstrapping per gene using the BCa method (Efron and Tibshirani 1994) implemented in the Boot package in R (Canty and Ripley 2012), and calculating X/A and Y/A on each iteration for 20000 replicates each. Bootstrapping was conducted on the final filtered set of 173 sex-linked, and 12355 autosomal genes. Note that the lack of recombination on the Y chromosome implies that statistical assumptions about independence across loci are violated, suggesting that the true uncertainty in the estimate Y/A may be wider than implied by bootstrapping results. To address this, we also used a maximum likelihood approach, implemented in a modified version of the MLHKA software (Wright and Charlesworth 2004), to independently estimate a credibility interval for the Y/A ratio (Figure S1). Because of the thousands of genes involved, a likelihood method incorporating divergence to control for heterogeneity in mutation rate was not feasible, as this would require maximizing the likelihood estimate of the mutation rate for each locus independently. Therefore, we assumed no heterogeneity in mutation rate, no recombination between Y-linked genes, and free recombination between autosomal loci. Our model thus had two parameters: $\theta_{autosomal}$ and Y/A. We varied both parameters and evaluated the likelihood for Y/A from 0.001 to 1, and $\theta_{autosomal}$ from 0.001 to 0.01.

### Neutral predictions and the effect of sex ratio bias on diversity

To test whether our estimated levels of diversity on X, Y and autosomal chromosomes could be explained by neutral processes, we compared our estimated levels diversity to predictions from a model for the expected $N_e$ of males and females when the variance in offspring number is smaller or larger than the standardly assumed Poisson (Kimura and Crow 1964; Hedrick 2011). As we were primarily interested in the influence of variance in male reproductive success on Y-linked $N_e$, and the parameter space within which this effect might generate Y/A and X/A ratios consistent with our data, we ignored non-poission variance in female offspring numbers, and calculated male $N_e$ as:

$$N_{e_m} = \frac{N_m k_m - 1}{k_m - 1 + \frac{V_{k_m}}{\bar{k}}} \quad (1)$$

where $k_m$ and $V_{k_m}$ are the mean and variance in male offspring number, respectively, in a population of $N_m$ reproducing males, and $\bar{k}$ is the total mean number of progeny (Kimura and Crow 1964). Given expected values of $N_{em}$ for a range of values for $V_{k_m}$ and $N_m$, the corresponding neutral expectation for autosomal and sex-linked $N_e$ at equilibrium were calculated as:

$$N_{e_A} = \frac{4 N_m N_f}{N_m + N_f} \quad (2)$$

$$N_{e_X} = \frac{9 N_m N_f}{4 N_m + 2 N_f} \quad (3)$$

$$N_{e_Y} = \frac{N_m}{2} \quad (4)$$

respectively, following (Wright 1931). Note that with equal sex ratios $N_f = N_m$, and Poisson-distributed offspring numbers

($V_k = \bar{k}$), then $N_{e_X}/N_{e_A} = 0.75$, and $N_{e_Y}/N_{e_A} = 0.25$ and in the standard model. Figure 1 shows these predictions as a function of the sex ratio for a range values of $V_{k_m}$, and we tested the fit of our observed Y/A and X/A diversity ratios to model predictions (shown in Figure 2) assuming the empirically-estimated *R. hastatulus* population sex ratio of 0.6 (Pickup and Barrett 2013).
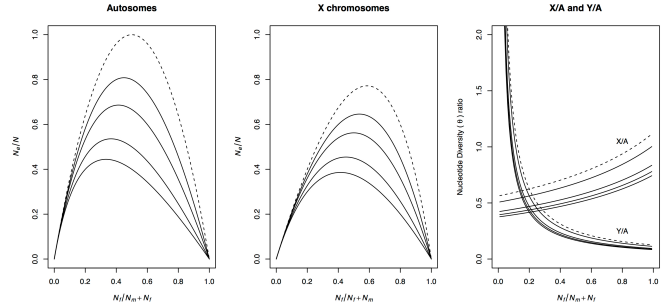


**Figure 1** The relation between effective population size and sex ratio bias for genes on autosomes (**A**) and the X chromosome (**B**), and corresponding normalized X/A and Y/A ratios (**C**). The sex ratio is shown as the proportion of females, $N_f/(N_f + N_m)$, plotted against $N_e/N$, where $N = N_m + N_f$. Dotted curves show equilibrium predictions when both males and females produce Poisson-distributed offspring numbers and ($V_k = \bar{k} = 2$). Solid curves correspond to increasing values for $V_{k_m}$, the variance in male offspring number, ranging from Poisson/2 (top solid curve in panels A and B) to 3*Poisson (bottom most curve). For (**C**) we assumed equal neutral mutation rates among genes and calculated the expected neutral diversity as $\theta = 4 N_e \mu$.

### Simulations of purifying selection

To study the effects of purifying selection on expected levels of Y-chromosome diversity, we conducted forward-time simulations of haploid Y chromosomes using the software SFSCODE (Hernandez 2008). We first estimated the distribution of fitness effects of deleterious mutations from our polymorphism data for X-linked genes using the method of (Keightley and Eyre-Walker 2007), which fits a gamma distribution of selection coefficients to the observed frequency distribution of nonsynonymous and synonymous polymorphism. We then used this estimated gamma distribution to parameterize the simulations, initializing each with our estimated $\theta$ from autosomal genes, but appropriately adjusted to reflect the expected reduction in $N_{eY}$ for a sex ratio of $N_f/(N_f + N_m) = 0.6$. To match our sample size and the number of synonymous sites sample from our data (see Supporting Information), the simulations sampled 6 haploid chromosomes, and contained 45,331bp of linked neutral sequence from which we calculated silent site diversity. To examine the expected reduction in diversity ($\pi/\pi_0$) as a function of the number of selected sites ($L$), we ran simulations over a range of values of $L$, up to a maximum of 5x10^6 (Figure 3A). To estimate $L$ from our data, we calculated the approximate likelihood of our observed data based on the proportion of simulations in which synonymous diversity, $\pi_{simulated}$, was not significantly different from our empirical estimate, $\pi_{observed}$ (Figure 3B).

## Results and Discussion

### Extensive loss of Y-chromosome diversity

Our analysis of polymorphism levels across the genome in the dioecious plant *R. hastatulus* reveals that neutral diversity on the Y chromosome is substantially lower than expected from a standard neutral model. In particular, neutral diversity on Y-linked genes $\theta_Y$ is on average ~2.1% of the value for their homologues on the X chromosome. Talking 1/4 of the mean autosomal diversity as the equilibrium expectation under neutrality ([Wright 1931](#)), this corresponds to a chromosome-wide reduction of ~93%. Note that by comparing X and Y diversity to autosomal diversity, our results indicate that the difference between X and Y homologues is not due to an elevation of diversity on the X chromosome, but rather a Y-specific reduction. Our results do suggest a slight elevation in X-linked diversity relative to the neutral prediction of X/A=0.75, but this difference is not significantly different from the expectation for the empirically-estimated ratio of 0.6. Although our estimates of diversity have not been normalized by divergence, previous work has shown that the average synonymous substitution rate, between *R. hastatulus* and the non-dioecious outgroup *R. bucephalophorus* is not significantly different for sex-linked (0.2016) and autosomal genes (0.219) ([Hough *et al.* 2014](#)). Furthermore, we found no evidence for significant differences in substitution rate between Y and X chromosomes ([Hough *et al.* 2014](#)). It is therefore unlikely that our results are caused by mutation rate differences between sex-linked and autosomal genes.

### Female biased sex ratios and variance in male offspring number

The occurrence of female-biased sex ratios in *R. hastatulus* is expected to lower Y diversity through a reduction in male $N_e$ and therefore the $N_e$ of the Y chromosome. In addition, male $N_e$ can be further reduced by high variance in male reproductive success (Figure 1), which is not unusual in annual wind-pollinated plants such as *R. hastatulus* that commonly exhibit extensive phenotypic plasticity in plant size and flower production (Harper 1977). Given that male plants in this species produce large amounts of pollen, and female flowers are uniovulate, there may be strong competition among males to fertilize females. Such competition should increase levels of X-linked diversity to a level that is close to (or even higher than) levels of autosomal diversity (Charlesworth 1996b), while simultaneously reducing Y-linked diversity. Thus, to test whether high variance in male reproductive success could explain our observed Y/A and X/A ratios, we tested the fit of our data to predictions from a model of variance in male reproductive success ([Kimura and Crow 1964](#)), using the estimated population sex ratio of 0.6 ([Pickup and Barrett 2013](#)). In common with most flowering plants, we do not have marker-based estimates of variance in male reproductive success in *R. hastatulus*. However, as shown in Figure 2, we find that, when evaluated at a sex ratio of 0.6, values for the variance in male reproductive success, $V_{km}$, that were large enough (range) to explain our observed reduction in Y/A diversity also predicted X/A diversity levels that were significantly greater than our observed values (range, P-value; Table 2). This indicates that the combined effects of a female-biased population sex ratio and extensive variance in male reproductive success is insufficient explain our observed reduction in Y-chromosome diversity (Figure 2; Table 2).

**Table 1** Observed vs. expected neutral diversity *R. hastatulus* sex chromosomes and autosomes.

| | Observed | | Expected | | | |
| | | | $V_{km} = 2$ | | $V_{km} = 4$ | |
| chr | $\theta$ | $/\theta_A$ | $\theta$ | $/\theta_A$ | $\theta$ | $/\theta_A$ |
|---|---|---|---|---|---|---|
| A | 0.006 | 1 | 0.006 | 1 | 0.006 | 1 |
| X | 0.0047 | 0.85 | 0.0019 | 0.33 | 0.0019 | 0.33 |
| Y | $10^{-4}$ | 0.002 | $10^{-4}$ | 0.002 | $10^{-4}$ | 0.002 |
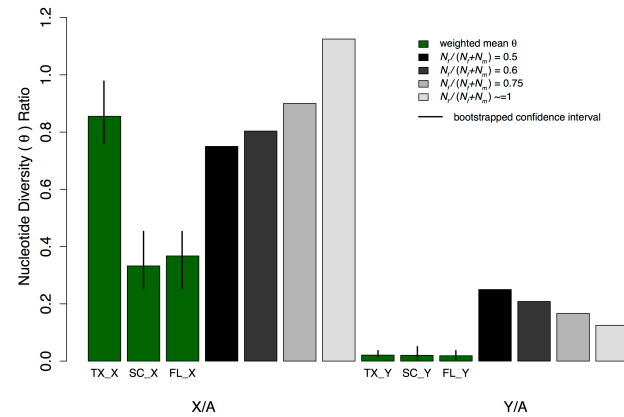


**Figure 2** new figure coming soon).

### Purifying selection

- Basic results; diversity loss consistent with purifying selection
  - HRI vs BGS; bgs over predicts diversity loss; good reasons to think we are in interference space. Include calculations from good et al; Figure 3 (sims); Figure 4 (bgs model -maybe put in supp)

Given that a large number of sites in the genome likely experience weak purifying selection (refs), the "interference" effect on Y chromosomes with suppressed recombination might alone cause a substantial diversity loss. Because there is no general analytical treatment describing the dynamics of such interference among a large number of linked sites under selection and drift, simulations are fundamental to investigating realistic situations. We therefore....

...Individual fitness effects may play a central role in single-site models, but we have shown that global properties like the variance in fitness and the corresponding linkage scale are more relevant for predicting evolution in interfering populations. "
- estimates of "Nes" ignore linkage by fiat under the assumption that sites evolve independently. But these estimates become unreliable precisely when small- and intermediate-effect mutations are most common - "Here, we have shown that simple behavior emerges in the limit of widespread interference. When fitness variation is composed of many individual mutations, the magnitudes and signs of their fitness effects are relatively unimportant. Instead, molecular evolution is controlled by the variance in fitness within the population over some effectively asexual segment of the genome" " In other words, we cannot conclude that interference is negligible just because Nes,

as inferred from data, is larger than one."

- Indeed, simulations suggest that the effects of HRI on 100,000 linked sites can reduce polymorphism levels by ~30 % compared to neutral expectations (McVean and Charlesworth 2000b), suggesting that HRI is likely important for the evolution of Y chromosomes.

### *Can selective sweeps of beneficial mutations explain Y-diversity loss?*

- Can we exclude sweeps empirically? Fitting frequency spectrum to model?; tajimas D? - Caveats; demography (e.g. changes in Ne) are not well known but could affect the metrics above; NeS assumes independence and can't be trusted; parameter uncertainties etc. - Implications for sex chromosome evolution in plants

I guess to my mind a sweep explanation is a priori implausible for the following reasons: 1) a species wide sweep, for example, would wipe out diversity entirely, so that the existing polymorphisms must have accumulated since the sweep, and that seems unlikely if none of our Y SNPs are singletons? 2. If recombination is suppressed within the region that we are working with (which it must be, because the SNPs segregate as fully sex-linked), then the occurrence of chromosome-wide linkage in that nonrecombining region will also cause beneficial mutations to have a very low probability of fixation. The relevant linkage scale here is the entire length of the non-recombining region, L. I've been doing a bunch of these HRI simulations for the last couple of days, and just as an example, in a simple model with the following parameters:

mutation rate: 1e-6 recombination rate: 0 generations: 1000 generations after equilibrium is reached (6000 generations total) L= 100000 (100kb) N=1000 individuals s =0.05 (beneficial)

the reduction in the probability of fixation of a beneficial mutation arising from linkage among 100,000 sites in a nonrecombining chromosome, is 98!!! (the reduction is relative to a model of independence among selected sites). That's huge (and it decreases with larger L). So at least on a first pass, the probability that a selected sweep is the explanation here seems very low.

### *Implications for the evolution of young Y chromosomes* Right.

On the other hand you can imagine that widespread silencing could happen precisely because of the problem of interference selection. For example, interference causes selection to be ineffective, so that Y chromosomes accumulate deleterious mutations and do not fix beneficial ones. This could then generate selection to down-regulate or silence such "unfit" Y-linked genes, as in an Orr Kim type of model. So yeah, I don't think that finding many selected sites would necessarily preclude a model of widespread gene silencing. It could just mean that we haven't gotten there yet (which would still be interesting).

## Conclusions

This is why I think HRI is so important for the Y. Not only does it result in a significant reduction in diversity when selection acts on linked deleterious mutations, but it also substantially reduces the probability of fixation of beneficial mutations. I grant that the probability of fixation of a beneficial mutation, and therefore the probability that a sweep explains our data, is not zero, but given the probable size of L on the R hastatulus Y chromosome, I don't think the claim that "our data is just as easily explained by sweeps" is defensible. If it is, then we haven't really done anything valuable here.

## Literature Cited

Bachtrog, D., 2008 The temporal dynamics of processes underlying y chromosome degeneration. Genetics **179**: 1513–1525.

Bachtrog, D., 2013 Y-chromosome evolution: emerging insights into processes of y-chromosome degeneration. Nature Reviews Genetics **14**: 113–124.

Bachtrog, D. and P. Andolfatto, 2006 Selection, recombination and demographic history in drosophila miranda. Genetics **174**: 2045–2059.

Bansal, V. and V. Bafna, 2008 Hapcut: an efficient and accurate algorithm for the haplotype assembly problem. Bioinformatics **24**: i153–i159.

Barton, N., 1995 A general model for the evolution of recombination. Genetical research **65**: 123–144.

Bergero, R., S. Qiu, and D. Charlesworth, 2015 Gene loss from a plant sex chromosome system. Current Biology **25**: 1234–1240.

Bull, J., 1983 *Evolution of sex determining mechanisms..* The Benjamin/Cummings Publishing Company, Inc.

Bustamante, C. D. and S. Ramachandran, 2009 Evaluating signatures of sex-specific processes in the human genome. Nature genetics **41**: 8–10.

Caballero, A., 1995 On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics **139**: 1007–1011.

Canty, A. and B. Ripley, 2012 boot: Bootstrap r (s-plus) functions. R package version **1**.

Charlesworth, B., 1978 Model for evolution of y chromosomes and dosage compensation. Proceedings of the National Academy of Sciences **75**: 5618–5622.

Charlesworth, B., 1991 The evolution of sex chromosomes. Science **251**: 1030–1033.

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

Charlesworth, B., 2001 The effect of life-history and mode of inheritance on neutral genetic variability. Genetical research **77**: 153–166.

Charlesworth, B. and D. Charlesworth, 2000 The degeneration of y chromosomes. Philosophical Transactions of the Royal Society of London B: Biological Sciences **355**: 1563–1572.

Charlesworth, D., 2015 Plant contributions to our understanding of sex chromosome evolution. New Phytologist **208**: 52–65.

Comeron, J. M. and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. Genetics **161**: 389–410.

Cotter, D. J., S. M. Brotman, and M. A. W. Sayres, 2016 Genetic diversity on the human x chromosome does not support a strict pseudoautosomal boundary. Genetics **203**: 485–492.

Efron, B. and R. J. Tibshirani, 1994 *An introduction to the bootstrap*. CRC press.

Ellegren, H., 2009 The different levels of genetic diversity in sex chromosomes and autosomes. Trends in Genetics **25**: 278–284.

Ellegren, H., 2011 Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nature Reviews Genetics **12**: 157–166.

Feldman, M. W., F. B. Christiansen, and L. D. Brooks, 1980 Evolution of recombination in a constant environment. Proceed-

ings of the National Academy of Sciences **77**: 4838–4841.

Field, D. L., M. Pickup, and S. C. Barrett, 2013 Comparative analyses of sex-ratio variation in dioecious flowering plants. Evolution **67**: 661–672.

Fisher, R. A., 1930 *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.

Good, B. H., A. M. Walczak, R. A. Neher, and M. M. Desai, 2014 Genetic diversity in the interference selection limit. PLoS Genet **10**: e1004222.

Grabowska-Joachimiak, A., A. Kula, T. Książczyk, J. Chojnicka, E. Sliwinska, and A. J. Joachimiak, 2015 Chromosome landmarks and autosome-sex chromosome translocations in rumex hastatulus, a plant with xx/xy1y2 sex chromosome system. Chromosome Research **23**: 187–197.

Hammer, M. F., F. L. Mendez, M. P. Cox, A. E. Woerner, and J. D. Wall, 2008 Sex-biased evolutionary forces shape genomic patterns of human diversity. PLoS Genet **4**: e1000202.

Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, M. P. Cox, and J. D. Wall, 2010 The ratio of human x chromosome to autosome diversity is positively correlated with genetic distance from genes. Nature genetics **42**: 830–831.

Hedrick, P. W., 2011 *Genetics of populations*. Jones & Bartlett Learning.

Hernandez, R. D., 2008 A flexible forward simulator for populations subject to selection and demography. Bioinformatics **24**: 2786–2787.

Hill, W. G. and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genetical research **8**: 269–294.

Hough, J., J. D. Hollister, W. Wang, S. C. Barrett, and S. I. Wright, 2014 Genetic degeneration of old and young y chromosomes in the flowering plant rumex hastatulus. Proceedings of the National Academy of Sciences **111**: 7713–7718.

Hough, J., S. Immler, S. C. Barrett, and S. P. Otto, 2013 Evolutionarily stable sex ratios and mutation load. Evolution **67**: 1915–1925.

Kaiser, V. B. and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends in Genetics **25**: 9–12.

Keightley, P. D. and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics **177**: 2251–2261.

Kimura, M. and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725–738.

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

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

Laporte, V. and B. Charlesworth, 2002 Effective population size and population subdivision in demographically structured populations. Genetics **162**: 501–519.

Li, H. and R. Durbin, 2010 Fast and accurate long-read alignment with burrows–wheeler transform. Bioinformatics **26**: 589–595.

Lunter, G. and M. Goodson, 2011 Stampy: a statistical algorithm for sensitive and fast mapping of illumina sequence reads. Genome research **21**: 936–939.

Mank, J. E., 2013 Sex chromosome dosage compensation: definitely not for everyone. Trends in genetics **29**: 677–683.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, *et al.*, 2010 The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. Genome research **20**: 1297–1303.

McVean, G. A. and B. Charlesworth, 2000a The effects of hill-robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155**: 929–944.

McVean, G. A. and B. Charlesworth, 2000b The effects of hill-robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155**: 929–944.

Muller, H. J., 1964 The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **1**: 2–9.

Muyle, A., N. Zemp, C. Deschamps, S. Mousset, A. Widmer, and G. A. Marais, 2012 Rapid de novo evolution of x chromosome dosage compensation in silene latifolia, a plant with young sex chromosomes. PLoS Biol **10**: e1001308.

Navajas-Pérez, R., R. de la Herrán, G. L. González, M. Jamilena, R. Lozano, C. R. Rejón, M. R. Rejón, and M. A. Garrido-Ramos, 2005 The evolution of reproductive systems and sex-determining mechanisms within rumex (polygonaceae) inferred from nuclear and chloroplastidial sequence data. Molecular Biology and Evolution **22**: 1929–1939.

Nomura, T., 2002 Effective size of populations with unequal sex ratio and variation in mating success. Journal of Animal Breeding and Genetics **119**: 297–310.

Ohno, S., 1967 *Sex chromosomes and sex-linked genes*, volume 1. Springer-Verlag.

Otto, S. P. and M. W. Feldman, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. Theoretical population biology **51**: 134–147.

Papadopulos, A. S., M. Chester, K. Ridout, and D. A. Filatov, 2015 Rapid y degeneration and dosage compensation in plant sex chromosomes. Proceedings of the National Academy of Sciences **112**: 13021–13026.

Pickup, M. and S. C. Barrett, 2013 The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. Ecology and evolution **3**: 629–639.

Pool, J. E. and R. Nielsen, 2007 Population size changes reshape genomic patterns of diversity. Evolution **61**: 3001–3006.

Quesada del Bosque, M., R. Navajas-Pérez, J. Panero, A. Fernández-González, and M. Garrido-Ramos, 2011 A satellite dna evolutionary analysis in the north american endemic dioecious plant rumex hastatulus (polygonaceae). Genome **54**: 253–260.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov, 2011 Integrative genomics viewer. Nature biotechnology **29**: 24–26.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay, D. Muzny, M. Platzer, G. R. Howell, C. Burrows, C. P. Bird, *et al.*, 2005 The dna sequence of the human x chromosome. Nature **434**: 325–337.

Smith, B. W., 1963 The mechanism of sex determination in rumex hastatulus. Genetics **48**: 1265.

Smith, B. W., 1964 The evolving karyotype of rumex hastatulus. Evolution pp. 93–104.

Watterson, G., 1975 On the number of segregating sites in genetical models without recombination. Theoretical population biology **7**: 256–276.

Westergaard, M., 1958 The mechanism of sex determination in dioecious flowering plants. Advances in genetics **9**: 217–281.

Wilson Sayres, M. A., K. E. Lohmueller, and R. Nielsen, 2014

Natural selection reduced diversity on human y chromosomes. PLoS Genet **10**: e1004064.

Wright, S., 1931 Evolution in mendelian populations. Genetics **16**: 97–159.

Wright, S. I. and B. Charlesworth, 2004 The hka test revisited. Genetics **168**: 1071–1076.

Zhou, Q. and D. Bachtrog, 2012 Chromosome-wide gene silencing initiates y degeneration in drosophila. Current Biology **22**: 522–525.