Hill-Robertson Interference Causes Reduced Genetic Diversity on a Young Plant Y-chromosome

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, natural selection is expected to eliminate neutral variation and lower 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 99.2% 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 evidence that this effect can substantially reduce nucleotide diversity on a young Y chromosome. Given the relatively recent origin of *R. hastatulus* sex chromosomes, 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. In both groups, for example, the loss of recombination between X and Y chromosomes is associated with an accumulation of deleterious mutations and a gradual loss of genes from the Y chromosome (Hough *et al.* 2014; Bergero *et al.* 2015; Bachtrog 2013). In some species, such genetic deterioration of the Y chromosome has also led to the evolution dosage compensation of the X chromosome (Charlesworth 1996b; Muyle *et al.* 2012; Mank 2013; Papadopulos *et al.* 2015). The independent evolution of these phenomena in such taxonomically distant groups as mammals and flowering plants suggests 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 theoretical and experimental biologists (Charlesworth 1978; Feldman *et al.* 1980; Barton 1995; Charlesworth 1996b; Otto and Feldman 1997; Charlesworth and Charlesworth 2000; McVean and Charlesworth 2000a).

One fundamental difference between X and Y chromosomes is that there are 1/3 as many Y-linked gene copies as X-linked ones in a diploid population. Therefore, genes on the Y chromosome are expected to experience an effective population size

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Saturday 3rd December, 2016%
[1]jhough@ucdavis.edu

(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 the equilibrium level of neutral polymorphism - proportional to the product of N_e and the neutral mutation rate, μ - 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; (Charlesworth *et al.* 1993)) and selective sweeps of strongly beneficial mutations (genetic hitchhiking; (Maynard Smith and Haigh 1974)). Furthermore, the build-up of genetic associations among selected mutations (high fitness covariance) on the Y means that selection will act non-independently across the chromosome, such that selection at any given focal site "interferes" with selection at the 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 equilibrium level of neutral variability (Fisher 1930; Muller 1964; Hill and Robertson 1966; McVean and Charlesworth 2000b; Kaiser and Charlesworth 2009; Good *et al.* 2014), with the magnitude of the effect depending strongly on the number and density of sites experiencing selection . These arguments all suggest that large non-recombining Y chromosomes should harbor a lower amount of neutral variability than predicted based on the number of Y chromosomes in a population (1/4 that of autosomes), and highlight that estimating the number of selected sites is a key test for determining the likely mechanism involved in driving diversity loss on a nonrecombining Y chromosome. However, the relative importance of neutral and selective factors for reducing chromosome-wide levels of diversity 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 on diversity of 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 silent site variability on X and Y chromosome, relative to values on autosomes. For example, high variance in male reproductive success reduces male N_e, which reduces the Y/A diversity ratio compared to the neutral expectation of 1/4, but it also increases 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) (but 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 are 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 selection (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), it is not clear whether purifying selection might have similarly 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)).

On the one 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, theory suggests that even weak purifying selection, if acting across a large number of genetically-linked 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 priori* likely to have strong effects on young plant sex chromosomes. On the other hand, if there has been widespread gene silencing during the early stages of Y degeneration, as found on the *Drosophila miranda* neo-Y chromosome (Zhou and Bachtrog 2012), then Y-linked diversity loss might be low in younger sex chromosome systems. In particular, if gene silencing occurs early during Y chromosome degeneration, few sites are expected to be experiencing purifying selection, with degeneration primarily driven by neutral genetic drift rather than inefficient selection. Distinguishing the relative roles of selective interference and neutral drift during Y chromosome evolution is a major open issue.

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 originated ~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, and to ensure that our diversity estimates are not biased by non-equilibrium conditions or other confounding factors arising from the X-autosome fusion in the XY_1XY_2 system, we focus only on the XY system.

Of particular importance for our study, *R. hastatulus* popula-

tions have 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 opportunity to test both the demographic and selective contributions to changes in sex-linked variability during sex chromosome evolution.

Materials and Methods

Population Samples and Sex-Linked Genes

We analyzed sex-linked and autosomal genes identified from Illumina RNA sequence data from 12 population samples (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 (Beaudry et al., in prep) 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 quality scores < 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 dupli-

cates (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 Wattersons (1975) estimator of the population parameter $\theta = 4N_e\mu$, where N_e is the effective population size, and μ 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 θ, 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 Y/A estimate may be wider than implied by bootstrapping. 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. However, previous analysis of divergence data does not suggest important chromosomal differences in synonymous divergence contributing to diversity heterogeneity (see Results). 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 f, the ratio of effective population size of the Y chromosome to autosomes. We varied both parameters and evaluated the likelihood for f from 0.001 to 1, and $\theta_{autosomal}$ per base pair from 0.001 to 0.01. In particular, we evaluated the likelihood of $\theta_{autosomal}$ and f, given the observed number of synonymous segregating sites for each i of n autosomal loci (S_i), the number of segregating sites on the Y chromosome (S_y), and the total number of sites per locus (L_i and L_y). The likelihood of ($\theta_{autosomal}$, f) is therefore given by:

$$L \propto P(S_Y, L_Y|f, \theta_a) \prod_{i=1}^{n} P(S_i, L_i|\theta_a) \quad (1)$$

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, including female-biased population sex ratios or high variance in male reproductive success, we compared our normalized X/A and Y/A diversity estimates to neutral predictions for a range of values for the ratio of effective population sizes of females and males. Because we were primarily interested in determining the parameter space within which a reduced male effective population size could result in Y/A and X/A diversity ratios consistent with our data, we tested the fit of our data to predictions across the full range of $N_{ef}/(N_{ef}+N_{em})$ (0.1 to 0.9; we excluded all-female or all-male cases). For autosomal and sex-linked effective population sizes given by:

$$N_{e_A} = \frac{4N_{em}N_{ef}}{N_{em}+N_{ef}} \quad (2)$$

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

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

(Wright 1931), the corresponding expected X/A and Y/A ratios an be expressed as:

$$\frac{N_{e_X}}{N_{e_A}} = \frac{9(-1+r)}{8(1-r)(-2+r))} \quad (5)$$

and

$$\frac{N_{e_Y}}{N_{e_A}} = \frac{1}{8r} \quad (6)$$

respectively, where $r = N_{ef}/(N_{ef}+N_{em})$. Note that when $N_{ef} = N_{em}$ ($r = 0.5$) then, $N_{e_X}/N_{e_A} = 0.75$, and $N_{e_Y}/N_{e_A} = 0.25$ as in the standard neutral model (Wright 1931). With a female-biased ratio, however, the expected X/A ratio can become greater than 1, approaching 1.125 in the limit as r$\rightarrow$ 1(*Caballero* 1995)(*Figure*1).

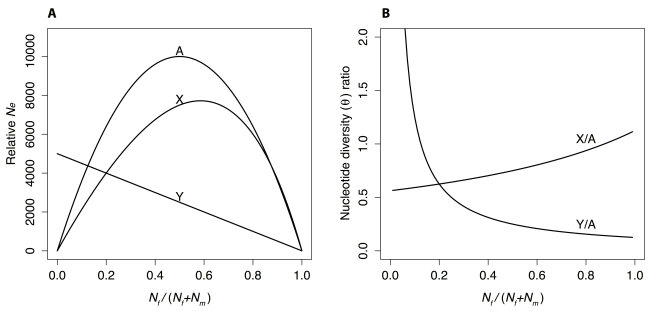

Figure 1 **A**. Expected effective population sizes, relative to N, for genes on autosomes (A), X chromosomes (X), and Y chromosomes (Y) as a function of $N_{ef}/(N_{ef}+N_{em})$. **B**. The corresponding X/A and Y/A ratios of diversity predicted at equilibrium (where $\theta = 4N_e\mu$), assuming equal neutral mutation rates for sex-linked and autosomal genes.

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 negative selection coefficients to the observed frequency distribution of nonsynonymous and synonymous polymorphisms. We then used this estimated gamma distribution to parameterize the simulations, initializing them with our estimated θ from autosomal genes, adjusted to reflect the expected neutral reduction in N_{eY}

for a sex ratio of $N_{ef}/(N_{ef}+N_{em}) = 0.6$. To match our sample size and the number of synonymous sites sampled from our data (see Supporting Information), the simulations sampled 6 haploid chromosomes, and the genome sequence contained 45,331bp of linked neutral sequence from which we calculated silent site diversity. To examine the expected reduction in diversity as a function of the number of selected sites ($L_{selected}$), we ran simulations over a range of values of $L_{selected}$, up to a maximum of $5x10^6$ (Figure 3A). To estimate $L_{selected}$, we calculated the approximate likelihood of our observed data based on the proportion of simulations in which synonymous diversity, $\pi_{simulated}$, was less than 0.0001 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 for Y-linked genes is on average ~2.1% of the value for their homologues on the X chromosome. Taking 1/4 of the mean autosomal diversity as the equilibrium expectation for θ_Y under neutrality (Wright 1931), this corresponds to a chromosome-wide reduction of 99.2% . Note that by normalizing X and Y diversity estimates by autosomal diversity levels, our results indicate that the difference between X and Y homologues is not due to an elevation of diversity on the X chromosome, but a Y-specific reduction. Interestingly, our results also suggest an elevation of X-linked diversity ($X/A = 0.85$) relative to the neutral prediction ($X/A = 0.75$), though the confidence intervals on this estimate are wide, ranging from 0.74 to 0.95 (Figure 2). Given the empirically-estimated sex ratio of 0.6 in *R. hastatulus*(Pickup and Barrett 2013), however, the elevation we observed is not unexpected; the neutrally-predicted X/A ratio for a sex ratio of 0.6 is $\approx$ 0.8 (Figure 2; Equation 4). 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), and 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.

Table 1 Observed and neutrally-expected silent site diversity on sex chromosomes and autosomes in *R. hastatulus*

	Observed		Expected	
chr	θ	$/\theta_A$	θ	$/\theta_A$
A	0.006	1	N/A	1
X	0.0047	0.85	0.0045	0.75
Y	10^{-4}	0.002	0.0015	0.25

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 thus a reduction the N_e of the Y chromosome. Male N_e could be further reduced by high variance in male reproductive success, 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. Given that male plants in this species produce large amounts of pollen, and female flowers are uniovulate, there may indeed 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 (Caballero 1995), while simultaneously reducing Y-linked diversity. In common with most flowering plants, we do not have marker-based estimates of variance in male reproductive success in *R. hastatulus*. We therefore tested whether an overall reduction in male N_e, arising either from high variance in male reproductive offspring number and/or a female biased population sex ratio, could explain our observed Y/A and X/A ratios, by comparing our data to neutral predictions across the full range of values for $N_{ef}/(N_{ef}+N_{em})$.

As shown in Figure 2, the expected reduction in the Y/A ratio across the full range of $N_{ef}/(N_{ef}+N_{em})$ is substantially higher than our observed reduction. Indeed, the lower limit for the Y/A ratio in a neutral model is 1/8, such that even in the extreme case where $N_{ef}/(N_{ef}+N_{em}) = 0.9$, the expectation for the Y/A ratio is $\approx$ 0.14, which is substantially higher than our observed estimate ($Y/A = 0.002$) (Table 1; Figure 2; Figure S3). Moreover, such large reductions in N_{em} also predict levels of X/A diversity that are not significantly larger than what we observed (Figure 2). Our results therefore indicate that, although reduced male N_e arising from sex-biased demography is expected to contribute to reduced Y chromosome polymorphism, it is insufficient to explain the Y/A reduction we observed.

Purifying selection

Selection against strongly deleterious mutations is expected be an important factor reducing genetic variability in regions with low recombination (Charlesworth *et al.* 1993; Charlesworth 1996a). The effects of such background selection (BGS) have been well studied theoretically (Charlesworth *et al.* 1997; Nordborg *et al.* 1996; Kim and Stephan 2000) and the theory has been used to explain patterns of genetic diversity across genomes (Comeron 2014) and plays a central role in explanations for the extensive diversity loss on Y chromosomes in mammals (Wilson Sayres *et al.* 2014) and Drosophila (McAllister and Charlesworth 1999; Charlesworth 1996b). Under background selection theory, the predicted loss of diversity is modeled as modeled as a reduction in the equilibrium N_e such that $\pi \approx 4f_0N_e\mu$, where $f_0 = e^{-\frac{U}{sh+r}}$, U is the mutation rate to strongly deleterious variants, s and h are selection and dominance coefficients, and r is the recombination rate. This yields the expected value of π relative to that in the absence of BGS (π_0): $E[\pi] = \pi_0 e^{\frac{-U}{2sh+r}}$, where , $\pi_0 = \theta = 4N_e\mu$. Because only reproducing males bearing Y chromosomes are relevant for determining the level of Y-linked diversity (i.e., $N_{eY} = N_e m/2$), the expected reduction in diversity is obtained in this haploid model by replacing N_e above with number of breeding males, N_{em}. Assuming that N_{em} is 0.4 of the total N_e estimated from

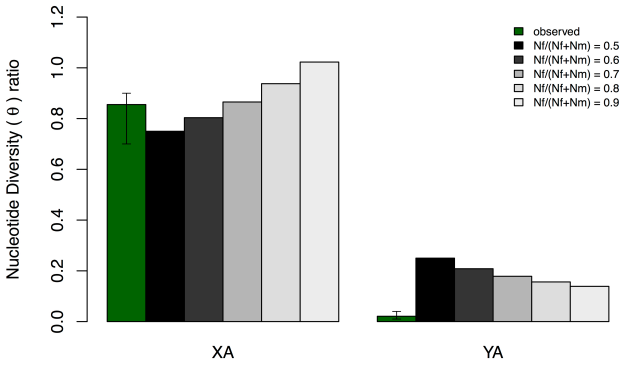

Figure 2 Observed and predicted X/A and Y/A ratios of neutral diversity. Predictions are shown for increasing values of N_{ef}/N_{em}. Observed estimates of X/A and Y/A were calculated as the average θ across genes, weighted by the number of synonymous sites per gene. Confidence intervals were calculated by bootstrapping (20000 replicates) using the BCa method (Efron 1987) implemented in the Boot package in R (Canty and Ripley 2012; R Core Team 2011), and using maximum likelihood for the Y-chromosome (see Methods).

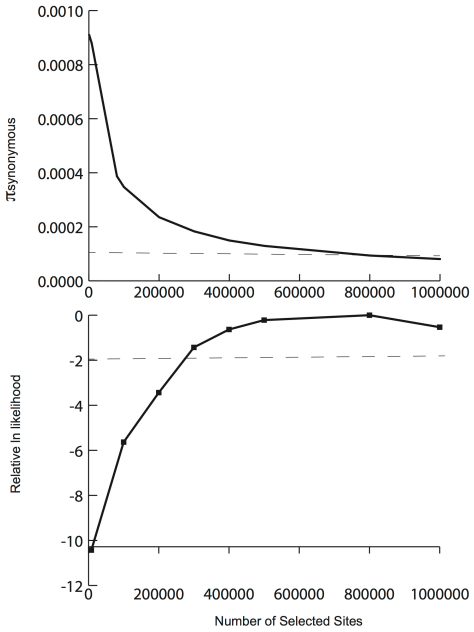

Figure 3 Simulation results of the level of neutral diversity pi_{syn} vs. the number of selected sites (top panel) with the dashed line corresponding to our point estimate for the level of neutral diversity on the Y chromosome, and the relative likelihood curve for the selected sites (bottom panel), with the dashed line corresponding to our approximate 95% credibility interval.

autosomes because of the female biased sex ratio (Pickup and Barrett 2013) and using s=0.003 as estimated from the distribution of fitness effects of deleterious mutations on the X chromosome, the corresponding expected value is $E[\pi] = 6x10^{-5}$ for a mutation rate of 10^{-8} per nucleotide per generation. Although there is considerable uncertainty in these parameters, the calculation yields a value of diversity that is 67% lower that what we observed (Table 1).

Importantly, BGS assumes independence among sites, and the theory breaks down when many linked sites are subject to selection at the same time, because the behavior of selected alleles interferes with the action of selection at linked sites (Good *et al.* 2014; Kaiser and Charlesworth 2009). As there are no analytical formulae for predicting the outcome of selection on many linked alleles experiencing selection and drift, however, simulations are fundamental to understanding this selective interference in realistic situations. We therefore conducted forward simulations of purifying selection using our estimated parameters for the distribution of fitness effects of deleterious mutations, and tested whether interference selection could result in a level of Y diversity at neutral sites similar to the level we observed. We found that models of purifying selection acting over a large number of linked selected sites resulted in an expected loss of diversity at neutral sites that was a remarkably good fit to our data (Figure 3). In addition, and consistent with previous studies, our simulations also suggest a saturating loss of diversity with an increasing number of selected sites (Kaiser and Charlesworth 2009).

Although we initialized our simulations with selection parameters estimated from the distribution of fitness effects of deleterious mutations, recent work has shown that, unlike with BGS, evolutionary dynamics in the interference selection regime are insensitive to differences in strength of selection, with expected diversity loss instead determined primarily by the population fitness variance over the non-recombining region, which is in turn expected to increase with the number of linked loci under selection (Good *et al.* 2014). Estimating the number of selected loci is therefore key to determining whether diversity on the Y chromosome may be affected by interference selection. Our maximum likelihood estimates of the number of selected sites, L, from our simulations, suggest that the number of sites subject to selection required to explain the diversity loss from the *R. hastatulus* Y chromosome is ($\approx 800kb$) (Figure 3).

Although our analyses of purifying selection models are consistent with our observed reduction in Y-linked diversity, it is also possible that selective sweeps have contributed. We have not formally excluded this possibility. However, if patterns of molecular evolution on the Y are driven primarily by interference selection, as our results suggest, then a further implication is that beneficial mutations should suffer a reduction in their probability of fixation, limiting the scope for sweeps to drive down diversity. Nonetheless, there is abundant evidence that soft sweeps are important determinants of patterns of nucleotide diversity (Messer and Petrov 2013), and it is possible that they have contributed to the extremely low levels of Y-linked variability that we observed.

Conclusions

The non-recombining region of a Y chromosome produces evolutionary dynamics that are similar to a haploid asexual population, with the lack of recombination resulting in build up of genetic associations among selected sites, a loss in the efficiency of selection, and an increase in population fitness variance (Fisher 1930; Muller 1964; Hill and Robertson 1966; McVean and Charlesworth 2000b; Kaiser and Charlesworth 2009; Good *et al.* 2014). Our study of neutral diversity levels on the relatively young *R. hastatulus* sex chromosomes provides clear evidence for an extensive loss of neutral diversity on the Y chromosome. Whereas neutral models of sex-biased demography were unable to explain the magnitude of diversity loss on this chromosome, forward population genetic simulations suggest that purifying selection acting over a large number of selected sites results in an expected level of Y-linked diversity that is consistent with what we observed. Our results are consistent with recent theory on interference selection, and suggest that interference is an important force in the evolution of young Y chromosomes. In particular, rather than widespread silencing of Y-linked genes and their subsequent degeneration through increased neutral drift, our study suggests that when a large number of linked sites are subject to purifying selection, Y-chromosome degeneration in the early stages may be largely driven by the effects of selective interference.

Acknowledgments

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., 1996a Background selection and patterns of genetic diversity in drosophila melanogaster. Genetical research **68**: 131–149.

Charlesworth, B., 1996b 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, B., M. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134**: 1289–1303.

Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical research **70**: 155–174.

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

Comeron, J. M., 2014 Background selection as baseline for nucleotide variation across the drosophila genome. PLoS Genet **10**: e1004434.

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. Proceedings 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.

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.

Kim, Y. and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155**: 1415–1427.

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.

Maynard Smith, J. and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genetical research **23**: 23–35.

McAllister, B. F. and B. Charlesworth, 1999 Reduced sequence variability on the neoy chromosome of drosophila americana americana. Genetics **153**: 221–233.

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.

Messer, P. W. and D. A. Petrov, 2013 Population genomics of rapid adaptation by soft selective sweeps. Trends in ecology & evolution **28**: 659–669.

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.

Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996 The effect of recombination on background selection. Genetical research **67**: 159–174.

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.