Purifying and Positive Selection Influence Patterns of Gene Loss and Gene Expression in the Evolution of a Plant Sex Chromosome System

Daisy Crowson

A thesis submitted in conformity with the requirements
for the degree of Master of Science

Department of Ecology and Evolutionary Biology
University of Toronto

Purifying and Positive Selection Influence Patterns of Gene Loss and Gene Expression in the Evolution of a Plant Sex Chromosome System

Daisy Crowson

Master of Science

Department of Ecology and Evolutionary Biology
University of Toronto

2016

Abstract

The degeneration of the Y chromosome is a hallmark of sex chromosome evolution. However, it is not known whether degeneration occurs as a direct result of selective interference due to lack of recombination, or whether adaptive gene silencing on the Y results in neutral degeneration. I used comparative transcriptome data from two related plant species with heteromorphic sex chromosomes, *Rumex rothschildianus* and *R. hastatulus,* to investigate processes underlying Y chromosome degeneration. The rate of degeneration varied greatly between the two species. In *R. rothschildianus,* I infer widespread gene loss. Genes showing lower constraint and those not expressed during the haploid phase were more likely to be lost. There was evidence of adaptive evolution on the Y from over-expression of Y alleles in certain genes. Targeted dosage compensation was occurring in more constrained genes. These results are consistent with selective interference playing a crucial role in the degeneration of the Y.

Acknowledgements

I would like to thank my co-supervisors Stephen Wright and Spencer Barrett for all their expert advice, encouragement and understanding. I would also like to thank Bill Cole, for his help in the greenhouse; Ramesh Arunkumar, for his help in the lab; Wei Wang, for his help with bioinformatics; Chris Balogh, for watering my plants while I was away; and David Timerman, who as well as taking care of my plants also helped keep me sane. Last, I would like to thank Stuart Campbell, who provided all of the mental, physical and emotional support that made this thesis possible. I also gratefully acknowledge a Connaught International Scholarship from the University of Toronto that funded me throughout the duration of this research. The research in this thesis was funded by NSERC Discovery Grants to Stephen I. Wright and Spencer C.H. Barrett.

Contents

List of figures

List of tables

Introduction

Heteromorphic sex chromosomes are a striking example of convergent evolution: they are found in a range of taxa as diverse as mammals, arthropods, algae, fish and flowering plants, and in some clades have arisen independently many times (Bull 1983; Bachtrog et al. 2011). This allows powerful comparisons to be made between closely related taxa in order to investigate the mechanisms governing sex chromosome evolution. The currently accepted model for the evolution of heteromorphic sex chromosomes involves accumulation of sexually antagonistic mutations (alleles that increase fitness in one sex but have a detrimental effect in the other) in the partially sex-linked pseudo-autosomal region. This process drives selection for suppression of recombination resulting in complete linkage between the sexually antagonistic allele and the sex-determination locus (Rice 1987a; Jordan and Charlesworth 2011). As well as a prerequisite for divergence between the X and Y chromosomes, loss of recombination is an important factor driving this divergence due to Hill-Robertson interference between genetically linked sites, where selection at one site interferes with selection at another linked site (Hill and Robertson 1966). This process is most easily understood in terms of a reduction in the local effective population size (N_e; Comeron et al. 2008; Charlesworth B and Charlesworth D 2010), caused by selection at one locus increasing variance in the 'reproductive success' of genetically linked sites. The overall effect is a loss of efficacy of selection, ultimately leading to genetic degeneration of the sex-limited sex chromosome (Y or W, but hereafter referred to as the Y chromosome for simplicity) due to accumulation of deleterious mutations, accumulation of repetitive DNA, loss of function of genes and eventually widespread gene loss (Charlesworth B and Charlesworth D 2000; Bachtrog 2013).

In contrast to model predictions, highly heteromorphic sex chromosomes with a fully degenerated Y are not the only observed outcome in the evolution of sex chromosomes. Some taxa have old sex chromosome systems that have not become

heteromorphic, but instead the X and the Y remain the same except for a small non-recombining region containing genes involved in sex determination (eg ratite birds and some snakes; Ogawa et al. 1998; Matsubara et al. 2006). This could be a result of an absence of sexually antagonistic mutations driving recombination suppression, or due to an alternative resolution of conflict involving sex-biased gene expression (Vicoso, Kaiser, et al. 2013). In other cases although recombination is suppressed and sex chromosomes become heteromorphic, the Y chromosome is not completely degenerated. A lack of extensive degeneration in sex chromosomes is particularly notable in flowering plants, where highly heteromorphic sex chromosomes are rare (Ming et al. 2011). Instead, sex chromosomes in plant species are often homomorphic or only moderately differentiated, with recombination occurring over most of their length, the case in papaya (VanBuren et al. 2015), *Populus* (Geraldes et al. 2015) and persimmon (Akagi et al. 2014). Moreover, two plant species with highly heteromorphic sex chromosomes in which Y chromosome degeneration has been investigated in depth (the unrelated *Rumex hastatulus* and *Silene latifolia*) reveal that although Y chromosome degeneration is occurring the extent of gene loss is comparatively low, estimated to be less than 30% based on transcriptome data (Hough et al. 2014; Bergero et al. 2015, respectively), although a recent study in *S. latifolia* using a genomic approach increased the estimate to 45% (Papadopulus et al. 2015). Although these cases are often viewed as an intermediate step in the evolution of a fully degenerate Y chromosome, it is also possible that partial degeneration is the long-term stable outcome of sex chromosome evolution in some taxa, just as homomorphic sex chromosomes are in other taxa.

There are several models describing how the non-independence of sites in non-recombining regions might drive genetic degeneration: background selection reduces the effective population size of linked regions due to the removal of continuously arising deleterious mutations (Charlesworth et al. 1993); Muller's ratchet drives stochastic fixation of deleterious mutations due to the loss of variants without them via genetic drift (Muller 1964); and selective sweeps reduce effective population size of linked regions due to fixation of beneficial mutations (Rice

1987b). These three processes are not mutually exclusive – on the contrary, they are closely related and are all likely to operate to some extent. Indeed, there is evidence for all three processes occurring on the Y chromosome (Bachtrog 2004; Kaiser and Charlesworth 2010; Singh et al. 2014).

The relative importance of the three non-independence models is predicted to vary between taxa, between sex chromosome systems and over time. Background selection and Muller's ratchet might be more important in the early stages of degeneration when the Y chromosome is still gene rich (Bachtrog 2008). Similarly, they might also be crucial in the degeneration of neo-Y chromosomes, as this involves a sudden cessation of recombination of many functional genes. On the other hand, selective sweeps might play a more important role at intermediate gene densities or in a case of sex-specific selection, when it is more likely that the selective advantage outweighs linked deleterious mutations (Bachtrog and Gordo 2004; Bachtrog 2008). Processes requiring strong selection (background selection and selective sweeps) might be particularly relevant to plants, as it has been estimated that ~2/3 of all genes are expressed in the male haploid gametophyte (pollen) (Honys and Twell 2004), which should lead to strong selection against the degeneration of these genes (Chibalina and Filatov 2011). However, background selection accelerates Muller's ratchet (Gordo and Charlesworth 2001), and therefore a large number of genes under strong purifying selection could result in an accelerated rate of degeneration of genes not expressed in the haploid phase. However, crucially, if these three models are the drivers of Y chromosome degeneration then it occurs despite a fitness cost. These models predict an inexorable accumulation of mildly deleterious missense and nonsense mutations at a rate that directly determines the rate of Y-chromosome degeneration. They also predict that degeneration should affect less important genes first, since these are more likely to be subject to mildly deleterious fitness effects.

While in the above models degeneration occurs primarily via inefficient selection on the non-recombining Y chromosome, it is also possible that most

genetic degeneration occurs neutrally following adaptive gene silencing. This could occur for two reasons. First, the Y chromosome is likely to experience a lower rate of positive selection compared with the X chromosome, which may select for a suppression of expression of the Y allele in favour of the better-adapted X allele (Orr and Kim 1998). Second, mildly deleterious mutations may accumulate on the Y due to inefficient selection, which may have strong enough fitness costs to actively select for suppression of expression (Zhou and Bachtrog 2012a). In both cases, following adaptive silencing the Y gene copy may degenerate entirely neutrally given its lack of expression.

There is some evidence for adaptive silencing from the preferential expression of X alleles over Y alleles in genes that still retain a Y copy, which has been found to occur in many taxa with partially degenerated Y chromosomes (e.g. Muyle et al. 2012; Hough et al. 2014; White et al. 2015). Furthermore, in the very young neo-Y chromosome of *Drosophila albomicans*, which displays very little sign of degeneration, Y alleles are already being expressed at a lower level than X alleles, suggesting silencing occurs before extensive degeneration (Zhou and Bachtrog 2012a). Distinguishing between these two neutral degeneration models requires looking at the type of genes that are lost first from the Y chromosome. If adaptive evolution on the X is the driving force, the prediction would be that genes that undergo more positive selection would be the first to be silenced and lost from the Y. In contrast, if mildly deleterious mutations due to inefficient selection on the Y are the driving force, the prediction might be that more highly expressed or more constrained genes should be the first to be silenced on the Y, due to the fitness costs of deleterious mutations on these being greater. Both these models predict that most of the Y chromosome degeneration occurs neutrally as a random accumulation of mutations in previously adaptively silenced genes.

Degeneration of the Y chromosome has genome wide consequences. The reduction in gene dosage caused by loss of a copy of the gene from the Y should select for the evolution of dosage compensation to restore optimal expression levels

within the genome (Vicoso and Bachtrog 2009). However, how dosage compensation occurs and to what extent varies hugely between different taxa (Mank 2013). In some taxa chromosome-wide changes have resulted in equal expression between the sexes, although exactly how this has occurred differs: for example, in Diptera the single X chromosome in males has up-regulated expression (Vicoso and Bachtrog 2015); in mammals one of the two X chromosomes is inactivated in females (Pessia et al. 2012); and in *C. elegans* both X chromosomes are down-regulated in females (Ercan et al. 2007). In placental mammals it has been shown that autosomal genes that interact with genes on the X chromosome are also down-regulated, explaining how within-genome optimal dosage is maintained (Julien et al. 2012). Other taxa also have global chromosome-wide changes, but fail to fully restore equal expression between the sexes (eg *Heliconius* butterflies, Walters et al. 2015). Finally, in many taxa there appears to be patchy dosage compensation, with particular genes having similar expression levels between males and females (eg Uebbing et al. 2012; Mahajan and Bachtrog 2015; Papadopulus et al. 2015).

However, evidence for patchy targeted dosage compensation is often unsatisfactory for two reasons. First, the loss of one copy of a gene sometimes results in an expression level that is greater than half the value provided by two copies, due to buffering provided by genomic network interactions (Malone et al. 2012). Some genes may therefore have similar expression levels without the need for any active dosage compensation. This is more likely to occur in genes that are expressed at low levels, which may explain why there are many reports of negative correlations between expression level and degree of dosage compensation (eg Melamed and Arnold 2007; Uebbing et al. 2012), despite the prediction that highly expressed genes should be preferential targets for dosage compensation due to their having greater fitness costs (Vicoso and Charlesworth 2009). Second, many reports of partial dosage compensation come from tests that are not necessarily appropriate for assessing differential expression, which requires an estimation of the variability in the data and associated error (Anders and Huber 2010). This may

explain the positive correlation occasionally observed between degree of compensation and between-individual variance in expression (e.g. Uebbing et al. 2012), which suggests the apparent dosage compensation is due to random experimental noise. It is therefore often unclear whether reports of patchy dosage compensation are truly the consequence of active selection for optimizing expression levels of particular genes.

Distinguishing between different models of Y chromosome degeneration and dosage compensation requires comparisons between sex chromosomes that vary in their rate and extent of degeneration in partially degenerated Y chromosomes. This situation is found in the large dioecious clade in the plant genus *Rumex* (Polygonaceae), which is divided into two smaller monophyletic clades with distinct sex chromosome systems. One clade has an XX/XY sex chromosome system and a Y-based sex-determining mechanism, whereas the second clade has an XX/ XY_1Y_2 sex chromosome system and a sex-determining system based on X/autosome balance (Smith 1969; Parker and Clark 1991). The second neo-Y chromosome in this latter clade may have arisen either via an ancient X-autosome fusion event, where the autosomal homologue in males formed the Y_2 chromosome, or a fission in the Y chromosome. The reduction in chromosome number in this clade (Smith 1969; Navajas-Perez et al. 2005) supports the former idea, although evidence from repetitive DNA accumulation is inconclusive (Steflova, Hobza, et al. 2013; Steflova, Tokan, et al. 2013). Nevertheless, this raises the possibility that the XX/ XY_1Y_2 has had a much larger non-recombining Y chromosome region over much of its history. Studies investigating the accumulation and distribution of repetitive DNA report a different degree of differentiation between sex chromosomes in these two groups: in the XX/XY_1Y_2 group the Y chromosomes are heterochromatic, with Y-specific satellite DNA families and recent expansions of repetitive DNA (Kejnovsky et al. 2013; Steflova, Tokan, et al. 2013). In the XX/XY group however, there appears to be less differentiation between sex chromosomes and less evidence of Y-specific repetitive DNA accumulation (Cuñado et al. 2007; Quesada del Bosque et al. 2011). If the sex chromosomes in these two clades have the same origin this suggests

genetic degeneration of the Y chromosome is occurring at different rates in the two sub-groups. However, it is also possible that the XY system is younger than the XY_1Y_2 system, either because sex chromosomes arose independently in the two lineages, or due to a turnover of sex chromosomes.

Here, I use comparative transcriptome data in two species of the genus *Rumex* to investigate the processes that govern sex chromosome evolution. Data from *R. rothschildianus*, a species belonging to the XY_1Y_2 clade with apparently highly degenerated Y chromosomes (Wilby and Parker 1988) is compared to *R. hastatulus* (see Hough et al. 2014), a species that belongs to the XX/XY clade but that has very recently acquired a neo-Y chromosome, also resulting in an XX/XY_1Y_2 system. I compare rates of degeneration between the two species, investigate patterns of gene loss in order to help distinguish between the different possible processes that may drive Y chromosome degeneration, and investigate evidence for targeted dosage compensation.

Results

Rate of gene loss varies between the two species

I used transcriptome data from two independent crosses of *R. rothschildianus* mapped to a reference transcriptome assembled *de novo* to identify genes located on the sex chromosomes from their diagnostic segregation patterns. From these segregation patterns I was able to distinguish genes that still have a Y allele expressed (hereafter 'sex-linked XY') from genes that have been lost or silenced on the Y chromosome (hereafter 'sex-linked hemizygous'). I investigated patterns and extent of Y chromosome degeneration in *R. rothschildianus*, and compared this to the previously identified sex-linked genes in *R. hastatulus,* using data from the North Carolina race with a very recent neo-Y chromosome (see Hough et al. 2014). The final number of sex-linked XY genes and sex-linked hemizygous genes identified in the two species are shown in **table 1**. *Rumex rothschildianus* has fewer identified sex-linked XY genes than *R. hastatulus*, but much higher levels of hemizygosity, estimated to be up to 92% of genes compared to an estimated 18% for *R. hastatulus*. Estimated gene loss is calculated using the number of sex-linked genes identified from X polymorphisms as opposed to X-Y differences, as the former are much rarer, making hemizygous genes less likely to be detected compared to sex-linked XY genes. The number of genes identified using different SNP cut-offs, and the number of autosomal genes identified are shown in **supplementary table S1**.

The difference in the extent of gene loss between the two species must be due either to a difference in the rate of Y chromosome degeneration or simply to a difference in the age of the sex chromosomes. Therefore, to estimate the age of the sex chromosomes and the pattern of recombination suppression, pairwise X-Y synonymous sequence divergence (d_S) was calculated for each gene (**fig 1**). Synonymous divergence shows a continuous distribution, with no evidence of distinct evolutionary strata, suggesting either recombination was suppressed

gradually or the stochastic nature of synonymous divergence is blurring the lines of distinct inversion boundaries. Median X-Y divergence in *R. rothschildianus* was 0.076, whereas in *R. hastatulus* it was 0.008. However, this difference is due to the large number of very young genes with very low divergence in *R. hastatulus,* as opposed to any difference in the range of d_S values (**fig 1**).

The age of sex chromosomes can be estimated using synonymous site divergence if a molecular clock is assumed. Although the highest d_S values should correspond to the oldest XY genes and therefore provide an estimate of when recombination was initially suppressed, synonymous site divergence is stochastic, and using extreme values in calculations based on average rates would introduce considerable bias. I therefore used the median value (plus and minus their associated error in the estimation; Yang and Nielsen 2000) of an arbitrarily defined old set of XY genes (d_S>0.1), which includes more than fifty genes in both species. This gave a median d_S value of old genes of 0.12-0.16 for *R. rothschildianus* and 0.13-0.22 for *R. hastatulus*. Dividing these values by half (as divergence is occurring along both the X and the Y branches) and using the direct estimate of spontaneous mutation rate in *Arabidopsis thaliana* of $7x10^{-9}$ per site per generation (Ossowski et al. 2010) gives an estimate of recombination suppression having occurred 8-11 million generations ago in *R. rothschildianus* and 9-16 million generations ago in *R. hastatulus.* Although there is considerable uncertainty associated with the calculations, these results do at least suggest the two sex chromosome systems are of similar age, which means the extent of gene loss in *R. rothschildianus* is more likely to be the result of a faster rate of Y chromosome degeneration compared to *R. hastatulus*.

The sex chromosomes are unlikely to be homologous between species

To be able to make appropriate comparisons between species about patterns of Y chromosome degeneration, it is essential to know whether the evolution of sex chromosomes has occurred independently or not. To determine whether the sex

chromosomes in the two *Rumex* clades are homologous I investigated the overlap between genes identified as sex-linked and autosomal in these two species. The number of orthologous genes that are sex-linked in both species, as well as the number of sex-linked genes in one species that were identified as autosomal in the other, are shown in **table 2**. Because both of these species have neo-Y chromosomes, even a single origin of sex chromosomes would only imply homology of the older sex chromosomes. To increase the power of the comparison I therefore also assessed the overlap of a subset of older sex-linked genes (with pairwise X-Y d_S > 0.1). The overlap between all sex-linked genes or a subset of old sex-linked genes between the two systems is not significantly different from the number expected by chance (permutation test, P=0.99) Moreover, the overlap between sex-linked genes in *R. rothschildianus* and autosomal genes in *R. hastatulus* is in fact greater than expected by chance (permutation test, $P<10^{-3}$), although the reciprocal comparison between sex-linked genes in *R. hastatulus* and autosomal genes in *R. rothschildianus* is not (permutation test, P=0.41). The percentage overlap of sex-linked or autosomal genes did not vary with the exact criteria used to classify genes into particular datasets, and there was no effect of number of SNPs used to define a gene as sex-linked on the percentage overlap observed, which would be expected if older sex-linked genes were shared between the species (**supplementary table S2**). Therefore, none of these comparisons provide evidence of homology between the sex chromosomes. Because old sex-linked genes are the most likely to be detected, due to a higher degree of X-Y divergence increasing the number of sex-linked SNPs, a low detection rate is unlikely to be causing this pattern. Overall, these data suggest these two *Rumex* clades have different sex chromosomes.

One factor that could affect rate of degeneration is the size of the non-recombining region, including sudden increases due to the formation of a neo-sex chromosome by the fusion of an autosome to a sex chromosome. To confirm the presence of *R. rothschildianus* in the XY_1Y_2 clade with an ancient neo-Y and highly heteromorphic sex chromosomes I built phylogenies using sequences from *R. acetosa* (known to be in the XY_1Y_2 clade), *R. hastatulus* (known to be in the XY clade)

and *R. bucephalophorus* (thought to be the a hermaphroditic out-group to these two clades; Navajas-Perez et al. 2005). I rooted the tree using *Fagopyrum esculentum,* a species in the same family as *Rumex,* as an out-group. As suggested by cytogenetic data (Wilby and Parker 1988), the consensus species phylogeny built from trees with highly (>80%) supported nodes clusters *R. rothschildianus* with *R. acetosa* in the XY_1Y_2 clade. However, the majority of gene trees do not place the hermaphrodite *R. bucephalophorus* as an outgroup to both clades with sex chromosomes (see **supplementary table S3**). However, all possible gene trees are well represented and wider taxon sampling would be required to resolve this node before a single or multiple origin of dioecy is confirmed. Taking the phylogenetic results and the lack of overlap between sex-linked genes identified in the two species together confirms *R. rothschildianus* is in a clade of species with an ancient neo-Y and highly heteromorphic sex chromosomes, and suggests that although a single origin of dioecy cannot be rejected the evolution of sex chromosomes has mostly occurred independently in these two clades, to the extent that there is no detectable homology between them.

Genes that are retained on the Y are also degenerating

Genetic degeneration is also likely to be occurring for genes still remaining on the Y chromosome, resulting in an accumulation of deleterious mutations even in genes that are still being expressed. I therefore further investigated Y-chromosome degeneration by comparing lineage-specific estimates of d_N/d_S between the X and the Y sequences for each species using various models of molecular evolution implemented in PAML. 'Branch models' allow d_N/d_S to vary on particular specified branches (here the X branch or the Y branch of either *R. rothschildianus* or *R. hastatulus*) compared to a background rate estimated for the rest of the tree, which also included the hermaphrodite *R. bucephalophorus*. These models were run separately for sex-linked genes in *R. rothschildianus* and *R. hastatulus*, as there were only 16 genes that were sex-linked XY in both species. The Y branch passed the likelihood ratio test for a significantly different rate of evolution to the background

rate more often than the X branch and it is more consistently at a higher rate in both *R. rothschildianus* and *R. hastatulus* (**table 3**). I estimated average branch-specific d_N/d_S using the 'free-ratios' model, which allows the rate to vary freely across the entire tree. Genes with very low X-Y synonymous divergence (d_S<0.001) were removed from the calculations of the mean, as they give highly skewed estimates of d_N/d_S. The average d_N/d_S was significantly higher on the Y branch compared to the X branch in both *R. rothschildianus* and *R. hastatulus* (Wilcoxon test, $P<10^{-12}$ for both comparisons) (**table 3**). The d_N/d_S ratio of the X branch for each gene in both *R. rothschildianus* and *R. hastatulus* was not significantly different from the rate of evolution of these same genes in the hermaphrodite *R. bucephalophorus*, where they are not sex-linked (Wilcoxon test, $P\geq0.54$ in both comparisons), whereas the Y is (Wilcoxon test, $P<10^{-15}$ in both comparisons), confirmation that the changes in d_N/d_S have occurred on the Y branch as opposed to the X branch.

An increase in d_N/d_S ratio can be driven by either a higher non-synonymous or lower synonymous rate of evolution. The synonymous site divergence (d_S) is also significantly different in *R. rothschildianus* between X and Y branches (mean d_s X = 0.037±0.002; mean d_s Y= 0.057±0.003; Wilcoxon test, $P<10^{-8}$), which implies either a higher mutation rate on the Y chromosome or reduced efficacy of selection on codon usage bias. However, this difference would result in a lower d_N/d_S ratio on the Y if it were the main driver of the difference in molecular evolution between the X and Y. Indeed, the difference in non-synonymous site divergence (d_N) is much larger (mean d_N X = 0.006±0.0005, mean d_N Y=0.021±0.002; Wilcoxon text, $P<10^{-15}$).

Reduced efficacy of selection due to loss of recombination is expected to result in the accumulation of mildly deleterious mutations, while strongly deleterious mutations are still purged (Charlesworth B and Charlesworth D 2000). This means loss of recombination should affect sites with lower constraint comparatively more. To investigate whether Y chromosome degeneration occurs at a faster rate in less constrained genes, I tested whether the difference between the rate of the X (which is affected by lower constraint of the gene) and the rate of the Y (which is affected

by lower constraint of the gene and reduced efficacy of selection) was greater in less constrained genes. There was a significant negative correlation between the difference between the evolutionary rates of the X and Y with both ancestral expression levels and the d_N/d_S ratio of the hermaphrodite relative (r_s= -0.18, P<0.05 and r_s= -0.22, P<0.02, respectively). This result confirms the prediction that loss of efficacy of selection is affecting weakly selected sites comparatively more, resulting in faster degeneration of less constrained genes.

Finally, I tested whether the Y chromosome was degenerating via frame-shift indels or nonsense mutations even in these genes that are still being expressed by comparing the length of the open reading frame (ORF) between X and Y transcripts. I found that the ORFs on the Y are significantly shorter than the X in both *R. rothschildianus* and *R. hastatulus* (Wilcoxon test, $P<10^{-7}$ in both comparisons). Overall, these results suggest genes that are still present on the Y chromosome are experiencing reduced efficacy of selection, and are experiencing degeneration due to the accumulation of deleterious non-synonymous missense and nonsense mutations in both species.

Degree of constraint of genes retained on the Y varies between the two species

The greater extent of gene loss in *R. rothschildianus* compared to *R. hastatulus* suggests an increased rate of Y chromosome degeneration in the former that may also be reflected in the genes that still retain expression of the Y allele. To investigate this I compared the rate of evolution of sex-linked XY genes between *R. rothschildianus* and *R. hastatulus*. There was no significant difference in either the X or the Y branch between these two species (Mann-Whitney U test, $P\geq0.1$ for both comparisons). However, I also calculated pairwise X-Y d_N/d_S, which was significantly higher in *R. hastatulus* compared to *R. rothschildianus* (median *R. rothschildianus* d_N/d_S= 0.322; median *R. hastatulus* d_N/d_S= 0.498; Mann-Whitney U test, $P<10^{-7}$) (**fig 2**). The lack of significance in the former comparison might be due to increased noise and reduced power when comparing lineage-specific as opposed

to pairwise differences, especially as most genes in *R. hastatulus* are very young, which might make lineage-specific estimates of d_N/d_S less reliable. The more robust pairwise comparisons suggest that, contrary to estimates of gene loss in the two systems, genetic degeneration of sex-linked XY genes in *R. hastatulus* is occurring at a faster rate than sex-linked XY genes in *R. rothschildianus.*

The sex chromosomes in these two species are not homologous, which means underlying differences in gene content might be driving the increased rate of non-synonymous evolution in *R. hastatulus* relative to *R. rothschildianus*. Moreover, the extensive gene loss from *R. rothschildianus* could mean sex-linked XY genes are no longer representative of the way most genes degenerated prior to being lost. To test whether there is some underlying difference in the sex-linked XY genes between the two species I compared the pairwise d_N/d_S ratio of the subset of 16 genes that were sex-linked XY in both species. There was no significant difference between the rate of X-Y divergence for these genes in the two species (Wilcoxon test, P=0.23), suggesting the overall pattern could be a result of the particular genes as opposed to a fundamental difference between the sex chromosomes. However, this could also be due to lack of power in a small set of genes. I investigated this further by comparing d_N/d_S ratio of the orthologs of sex-linked XY genes in the out-group hermaphrodite species *R. bucephalophorus*. In *R. bucephalophorus* genes that are sex-linked XY in *R. rothschildianus* were significantly more constrained than genes that are sex-linked XY in *R. hastatulus* (median d_N/d_S 0.16 and 0.20, Mann-Whitney U test, P<0.02). When this same comparison is done for all sex-linked genes, including genes that have been lost for the Y chromosome in both species, the difference between the two sets of genes disappears (median d_N/d_S 0.18 and 0.20, P=0.67). This is consistent with the idea that there are differences in the particular genes that are sex-linked XY in the two species, suggesting that widespread gene loss is changing the composition of sex-linked XY genes in *R. rothschildianus,* resulting in them being on average more constrained (see below).

Finally, I also exploited the negative correlation between X-Y divergence and ancestral gene expression (**supplementary fig 1)** in both *R. rothschildianus* (r_s= -0.22, P<0.001) and *R. hastatulus* (r_s= -0.17, P<10^{-6}), which suggests highly expressed genes are under more selective constraint than genes expressed at lower levels, to perform a second comparison of constraint between the two sets of genes. I compared average ancestral expression, and found that it was lower for genes that are sex-linked XY in *R. hastatulus* than *R. rothschildianus* (Mann-Whitney U test, P<0.003), again suggesting these genes are under less selective constraint, explaining the greater pairwise X-Y d_N/d_S in *R. hastatulus* compared to *R. rothschildianus.*

Gene loss from the Y chromosome is not random

The large number of genes in both sex-linked XY and sex-linked hemizygous datasets in *R. rothschildianus* allows investigation into the processes governing gene loss from the Y – in particular, what kinds of genes are lost, and what implications this has about the various models describing Y chromosome degeneration. Furthermore, the lack of overlap in sex-linkage between *R. rothschildianus* and *R. hastatulus* allows us to examine ancestral patterns of constraint and gene expression of these genes prior to them becoming sex-linked. The rate of molecular evolution of hemizygous genes in *R. rothschildianus* (**table 4**) was significantly higher than autosomal genes (Mann-Whitney U, P<0.01), although not significantly different from genes on the X chromosome that still retain a Y copy (Mann-Whitney U, P=0.12).

To investigate whether the lower constraint was a property of the genes themselves or a result of their hemizygosity (for example, due to decreased effective population size) I compared rates of evolution of genes orthologous to hemizygous genes in *R. rothschildianus* to genes that are orthologs to autosomal genes and sex-linked genes retained on the Y. Orthologs of hemizygous genes in *R. rothschildianus* have a significantly faster rate of evolution than the other classes in both *R.*

bucephalophorus (Mann-Whitney U, $P<0.05$ for both comparisons) and *R. hastatulus* (Mann-Whitney U, $P<0.02$ for both comparisons), suggesting the lower constraint is a cause rather than a consequence of hemizygosity. Although the difference between sex-linked hemizygous and sex-linked genes that retain a copy on the Y chromosome in *R. rothschildianus* is not significant, this could be due to lower effective population size of the X leading reduced efficacy of purifying selection. To further compare the likely degree of constraint experienced by genes that have been lost compared to genes that retain a Y copy, I compared ancestral expression of these genes. The ancestral expression of hemizygous genes is significantly lower than for genes that remain on the Y (median normalized ancestral expression was 614.5 for sex-linked XY genes and 456.3 for sex-linked hemizygous genes, Mann-Whitney U test, $P<0.02$). Overall, the results suggest that less constrained genes are more likely to be lost from the Y chromosome.

However, it is also possible that the higher rate of molecular evolution is due to prevalent positive selection in hemizygous genes, either as a cause of their hemizygosity, under a model of adaptation on the X driving silencing of the Y, or a consequence of hemizygosity, due to more efficient positive selection of recessive alleles in hemizygous genes (Faster-X effect; Charlesworth et al. 1987). To investigate this I fitted a branch-site model in PAML, allowing rates to vary between branches as well as between sites, which provides a test for positive selection ($d_N/d_S>1$) operating on particular sites in particular lineages. The number of genes passing the LRT for the presence of positive selection in hemizygous genes was lower in *R. rothschildianus* (18%) than for the same genes in *R. hastatulus* (29%). Moreover, the percentage of genes under positive selection is higher in autosomal genes (26% in *R. rothschildianus*; 39% in *R. hastatulus*) and in sex-linked XY genes (22% in *R. rothschildianus*; 21% in *R. hastatulus*) than in genes that have become hemizygous in *R. rothschildianus*. This suggests that widespread positive selection is neither a cause nor a consequence of hemizygosity, and instead the higher average d_N/d_S is a result of lower selective constraint.

Genetic degeneration of the Y chromosome in plants may be limited by extensive gene expression in the male gametophyte. To test whether selection during the haploid pollen phase of the life cycle limits the extent of gene loss from the Y chromosome I used pollen and sperm transcriptome data from *Arabidopsis thaliana* and *Nicotiana tabacum* to identify genes expressed in the haploid phase of the lifecycle, and tested for an underrepresentation of haploid-expressed genes in hemizygous genes compared to genes that remain on the Y (**table 5**). There was a significant underrepresentation of haploid-expressed genes in the set of genes that have been lost from the Y chromosome (Fisher's exact test, $P<0.001$), whereas the difference between all sex-linked genes and autosomal genes was not significant (Fisher's exact test, $P=0.67$). This suggests genes expressed during the haploid phase of the life cycle are less likely to be lost from the Y chromosome than genes that are only expressed in diploid tissue.

A source of bias would arise if genes expressed in haploid phase are on average more constrained than genes that are only expressed in the sporophyte, as I have already shown less constrained genes are more likely to be lost. Indeed, in *R. rothschildianus* genes expressed in the haploid phase are significantly more constrained than genes not expressed in the haploid phase (median 0.160 and 0.168 respectively, Mann-Whitney U test, $P<0.03$). However, the difference is small, and the same genes are not significantly different from each other in the hermaphrodite out-group *R. bucephalophorus* (Mann-Whitney U test, $P=0.34$). Therefore, it seems unlikely that this is the only factor driving the pattern of gene loss, but rather that haploid expression affects rate of gene loss over and above degree of constraint.

Finally, I carried out a functional enrichment analysis of genes that remain on the Y chromosome compared to genes that are lost from the Y chromosome, to see whether some types of genes are more likely to be lost than others. Sex-linked XY genes were enriched for various fundamental categories, such as intracellular transport, protein transport, post-embryonic development, and RNA binding, splicing and processing (**supplementary table S4**), whereas sex-linked hemizygous

genes were enriched for fewer categories of likely less fundamental processes (**supplementary table S5**).

Patterns of gene expression change in response to gene loss

Y chromosome degeneration is in some species accompanied by dosage compensation to regain optimal levels of gene expression within the genome despite the loss of one copy of the gene. To investigate whether *R. rothschildianus* shows evidence of dosage compensation I investigated expression levels of hemizygous genes between male and female individuals (**fig 3A**). Hemizygous genes in *R. rothschildianus* have a male over female expression ratio significantly above 0.5, although it does not reach the 1:1 ratio expected if full dosage compensation has occurred. Ancestral gene expression (estimated using *R. hastatulus*, where the genes are not hemizygous) shows that biased expression is unlikely to be ancestral and is instead a consequence of hemizygosity (**fig 3A**). To test whether the change has occurred in males or in females I investigate the ratio of female/ancestral and male/ancestral expression (**fig 3C**). The ratio of present over ancestral expression is shifted above the 0.5 line for males, whereas female expression is on the 1:1 line, indicating the changes in expression have primarily occurred in males, as a result of the loss of one copy of these genes. The same tests were carried out on autosomal genes to validate the approach of using *R. hastatulus* as an estimate of ancestral expression, and in each case the present over ancestral ratio was 1:1 (**fig 3B** and **fig 3D**). These results suggest that despite widespread gene loss complete dosage compensation has not developed in *R. rothschildianus*.

It is possible that particular genes have been targeted for dosage compensation, as opposed to the whole chromosome. This is perhaps more likely given that *R. rothschildianus* still has many genes present on the Y chromosome that would not require correction of dosage. To investigate this I looked at the male/female ratio on a gene-by-gene basis (**fig 4**). In a scenario of full dosage compensation most genes should fall on the 1:1 line, whereas under a scenario of no

dosage compensation they should fall close to the 0.5:1 line. Although in some cases hemizygous gene expression in males is close to the null expectation of half the expression level found in females, there are many genes that are above this line. However, between-individual variance introduces noise in these estimation that make it hard to determine exactly when a gene is likely to have been "compensated" and when it is just random noise. I therefore directly tested for statistically significant differential expression (see Methods) to more robustly categorise these genes into "dosage compensated" or "not dosage compensated" depending on whether males had expression levels that were significantly higher than half the value of female expression (**table 6**). Approximately ~30% of hemizygous genes are expressed at significantly higher levels than half the value of female expression, and a large subset of these (about ~25% of all hemizygous genes) do not have significantly different expression between males and females.

To investigate whether genes defined as "dosage compensated" were likely to be actively targeted for dosage compensation I compared the male/female expression ratio to the degree of ancestral constraint of hemizygous genes, using this as a proxy for its overall importance. Because genes expressed at higher levels might be harder to achieve dosage compensation for, and more constrained genes have higher expression (**supplementary figures S2 and S3**), I included ancestral gene expression in a partial correlation test of degree of compensation and d_N/d_S ratio, which was significant (r_s= -0.30, P<0.05) (**fig 5**). This suggests that targeted dosage compensation is occurring, up-regulating the expression of more selectively constrained, and therefore presumably more important, genes in males. I carried out two further tests to confirm these genes are likely to be targeted for dosage compensation. First, I compared their ancestral expression levels, because genes with low expression levels are more likely to have buffered expression levels resulting in "automatic" dosage compensation (Malone et al. 2012). However, there was no significant difference between ancestral expression levels between genes defined as "dosage compensated" compared to genes defined as "not dosage compensated" (Mann-Whitney U test, P=0.92) or a significant correlation between

male/female expression ratio and ancestral gene expression (r_s= 0.1, *P*=0.47). It is also clear from **fig 4** that not all genes with male/female ratio close to 1 are expressed at low levels. Second, I tested for a correlation between male/female expression ratio and the between-individual variance, which if significant would suggest apparent dosage compensation was a statistical artefact. However, there was no significant correlation between degree of dosage compensation and between-individual variance (r_s= -0.02, *P*=0.89). Overall, these results suggest that about a quarter of hemizygous genes in *R. rothschildianus* show evidence of targeted dosage compensation.

Because changes in expression can occur in both males and females, I used ancestral expression levels to determine the direction of change in gene expression. First, I asked whether genes in which male expression is not significantly different from half the value of females (for which therefore apparently no dosage compensation has occurred) are expressed at ancestral levels in females, or whether some up-regulation of male transcription has led to similar increased female gene expression, keeping the male/female expression close to 0.5 but shifting the female expression away from optimal. The ratio of female over ancestral expression is shifted slightly above one, whereas male expression over ancestral is just over 0.5 (**fig 6A)**. However, the overlapping notches suggest this difference is not significant. Moreover, a more rigorous test for differential expression between current and ancestral levels showed that current female expression varied equally in both directions relative to ancestral levels (35 genes significant, 40% of them lower than ancestral) whereas male expression was biased towards being significantly lower than ancestral (44 genes significant, 75% of them lower than ancestral). These results suggest that the lower expression levels in males are not ancestral and females have not moved away from optimal expression. Second, I asked whether genes that have been compensated have simply been up-regulated in males, resulting in equal gene expression compared to ancestral expression levels. **Fig 6B** suggests current expression levels of genes that have been dosage compensated are below ancestral levels. Supporting this, compensated genes that show significantly

different expression between current and ancestral levels are biased towards lower current expression in both females (16 genes significant, 11 of these lower than ancestral) and males (14 genes significant, 10 of these lower than ancestral). This suggests that compensation has not occurred entirely through up-regulation of male transcription, but also through down-regulation of female expression.

The two models that describe a more neutral Y chromosome degeneration predict that expression of Y alleles should be actively suppressed relative to X alleles due either to adaptive evolution on the X or slightly deleterious mutations on the Y. To test for this I measured allele-specific expression for X and the Y alleles in males, as well as reference and alternate alleles in sex-linked genes in females and in autosomal genes (**fig 7**). Whereas the latter two overlap the 1:1 line, in males the X is expressed at a higher level than the Y. If this was due to an active suppression of the Y due to either of the two models outlined above, the Y/X expression is expected to be lower in genes with higher X-Y d_N/d_S (where more deleterious mutations have accumulated) or in genes that showed evidence of positive selection occurring on the X chromosome. However, there was no correlation between Y/X expression ratio and X-Y d_N/d_S (r_s= 0.01, *P*=0.86) and no difference between Y/X expression ratio between genes that showed evidence of positive selection on the X chromosome compared to genes that did not (based on results from PAML branch-site model likelihood ratio test for positive selection; Mann Whitney U, *P*=0.96). Additionally, there is no correlation between Y/X expression and time since recombination suppression, estimated by X-Y synonymous divergence (r_s= -0.05, *P*=0.41). These results suggest that the lower expression of the Y allele relative to the X is more likely to be a passive consequence of Y chromosome degeneration than a consequence of adaptive active suppression of Y chromosome expression.

A reduction in expression of Y alleles could lead to lower overall expression of sex-linked XY genes in males if it was not accompanied by an equal increase in expression of the X allele. To test for this I looked at overall expression differences between males and females in XY genes (**table 6**). Sex-linked genes have a much

higher percentage of genes expressed significantly differently between the sexes compared to autosomal genes; however, differentially expressed genes are biased towards being higher in males (**table 6** and **supplementary fig S4**). Moreover, the increase in gene expression in males is not arising from expression of a higher-quality X allele, but is instead the result of over-expression of the Y allele compared to the X allele (**fig 8**). This suggests that this difference in levels of gene expression between the sexes may be a result of male-specific adaptive Y alleles. To further investigate this I conducted a functional enrichment analysis of genes with a Y-allele expression bias, and found that they were enriched for categories of genes that are likely to have different optima between the two sexes, such as negative regulation of flower development and reproductive structure development (**supplementary table S6**). These could have been sexually antagonistic alleles present in the pseudo-autosomal region before recombination suppression resulted in perfect linkage, and therefore are not necessarily newly arisen variants on the Y. However, if the bias in expression is due to an the Y allele being beneficial in males then positive selection must have at least occurred on Y-specific expression levels, suggesting that although the Y chromosome is highly degenerate in *R. rothschildianus*, male-driven adaptation might still be occurring.

Discussion

Variation in the rate and extent of Y chromosome degeneration

The estimate of the percentage of genes lost or silenced on the Y chromosome is very different for *Rumex rothschildianus* (over 90%) compared to *R. hastatulus* (less than 20%) (**table 1**). Estimated gene loss in *R. hastatulus* will be affected by the very recent acquisition of a neo-Y chromosome, which inflates the number of sex-linked XY genes while presumably not immediately increasing the number of sex-linked hemizygous genes. However, estimates from the *R. hastatulus* race without the neo-Y chromosome are still below 30% (Hough et al. 2014). Although there must be considerable uncertainty in the estimation of the age of sex chromosomes, which makes it hard to compare rates of degeneration between unrelated taxa, the estimates made here that sex chromosomes in *R. rothschildianus* and *R. hastatulus* are of similar age is more robust, even if the precise age is not known. Genes are therefore being lost from the Y chromosome in *R. rothschildianus* three times faster than from the Y chromosome in *R. hastatulus*.

The rate of molecular evolution of sex-linked XY genes as measured by pairwise X-Y d_N/d_S divergence is higher in *R. hastatulus* than *R. rothschildianus* (**fig 2; table 3**), suggesting that there is no clear difference in the amount of degeneration that is occurring or will occur, but instead there is only a difference in the amount that has already occurred. That is, *R. rothschildianus* has lost its less constrained genes from the Y chromosome, whereas in *R. hastatulus* these are still present but are accumulating a high frequency of deleterious mutations. Although it is clear from other comparative studies that the rate and extent of the degeneration of the sex-limited sex chromosome varies greatly (eg Vicoso, Emerson, et al. 2013; Zhou et al. 2014), these comparisons often highlight disparities between very different taxa, including major differences in the age of sex chromosomes, the sex-chromosome system and fundamental aspects of life history. The difference shown

here between *R. rothschildianus* and *R. hastatulus* demonstrates that even closely related taxa with sex chromosomes of roughly similar ages can also have extremely different rates of degeneration.

There are several possible explanations for the difference in degeneration rate between *R. rothschildianus* and *R. hastatulus*. First, although the origin of the sex chromosomes occurred at a broadly similar time, the neo-Y chromosome in *R. rothschildianus* is much older, as it is shared with the other species in the clade (Navajas-Perez et al. 2005). Neo-sex chromosomes might be important in accelerating degeneration due to the sudden cessation of recombination in a large gene-rich region, which will greatly increase the effects of both background selection and Muller's ratchet (Bachtrog 2013). This could result in the increased genetic degeneration in *R. rothschildianus* compared to *R. hastatulus*, which only acquired its neo-Y chromosome relatively recently (Hough et al. 2014). Second, differences in ancestral population size could also affect Y chromosome degeneration, as both background selection and Muller's ratchet are accelerated with smaller effective population sizes (Charlesworth B and Charlesworth D 2000). Third, differences in rates of positive selection between the two species could also contribute to different rates of degeneration, either due to selective sweeps on the Y chromosome, or through silencing of the Y due to a better-adapted X chromosome. However, why rates of positive selection should vary this much between these two taxa is unclear. Last, because these chromosomes are unlikely to be homologous (**table 2**), it is possible that the particular suite of genes present on the proto-Y chromosomes affected how quickly they degenerated. However, there was no difference in constraint between the two sets of sex-linked genes (including both lost and retained genes) in the hermaphrodite relative *R. bucephalophorus*, therefore this scenario is unlikely. Although the effect of the neo-Y chromosome is arguably the most plausible explanation, whether it can alone account for such a large difference in rate of gene loss is debatable. With only two species the definitive reasons behind the striking differences between rate and extent of genetic degeneration of their Y chromosomes cannot be determined – however, these

results confirm the value of comparative approaches for addressing outstanding questions in sex chromosome evolution, and suggests plant genera with multiple species with sex chromosomes, like *Rumex*, are a rich source for future comparative transcriptomic and genomic studies.

The absence of any plant species with completely degenerated Y chromosome and scarcity of heteromorphic sex chromosomes in general, has led to the idea that the Y chromosome does not degenerate as much or as quickly in plants. The reason invoked is selection occurring during the haploid phase of the lifecycle, as ~2/3 of the genome is expressed in male haploid gametophyte (Honys and Twell 2004), which could mean that gene loss is only tolerated for about a third of genes on the Y chromosome. The extent of gene loss or silencing reported here for *R. rothschildianus* is greater than for any other plant that has been investigated previously. The estimate of over 90% of genes (**table 1**) being absent from the Y chromosome is 2 or 3 times higher than estimates from *S. latifolia* and *R. hastatulus* (Hough et al. 2014; Papadopulus et al. 2015).

There are several possible explanations for the much higher extent of gene loss in *R. rothschildianus* than the predicted maximum 1/3 of all genes. First, it is possible that not all of the genes expressed in the male gametophyte are fundamental in its function, thus freeing up a larger proportion of the genome to degeneration. Second, an unlikely but not impossible scenario is that the proto-sex chromosomes happened to have a very high proportion of the genes that were not expressed in the haploid phase of the lifecycle, and could therefore be lost from the Y chromosome. Third, it is possible that these genes are switched off in diploid tissue but are still present in the genome and can be expressed in the haploid phase. Last, it is possible that the reduced efficacy of selection due to loss of recombination is strong enough that degeneration occurs despite strong selection against it, resulting in important genes being lost. This would result in Y-bearing pollen being less competitive than X-bearing pollen, a phenomenon called certation (Correns 1928). There is indirect evidence that this might occur from the widespread female-

bias in sex ratio in plants with sex chromosomes (Field et al. 2012a), which in *Rumex* has been shown to at least partly be determined at a pre-zygotic level (Stehlik and Barrett 2006; Blocka-Wanda et al. 2007; Field et al. 2012b). This could result in an interesting balance between selection for equal sex ratios (Fisher 1930) and the degeneration of the Y chromosome, potentially resulting in stable female-biased sex ratios (Hough et al. 2013).

It is important to note that transcriptome data cannot distinguish between genes that have been deleted from the Y from those that are still present but for which expression has been suppressed. However, two studies explicitly comparing lack of expression with DNA sequence in *S. latifolia* concluded that in most cases lack of expression of the Y allele was due to losses of the genes themselves, being either entirely absent or interrupted by premature stop codons (Bergero et al. 2015; Papadopulus et al. 2015). These studies also resulted in a higher estimate of gene loss than when solely based on transcriptome data, suggesting our methods are conservative. Similarly, genomic evidence to date suggests that in *R. hastatulus* the majority of genes identified as hemizygous using transcriptome data are deleted from the genome (Beaudry et al. in preparation). However, it has also been suggested that gene silencing is one of the earliest signs of Y chromosome degeneration, preceding any gene loss (Zhou and Bachtrog 2012a). Cytological studies have shown that *R. rothschildianus* has an almost entirely heterochromatic Y chromosome (Wilby and Parker 1988), and therefore it is possible that wide-scale gene silencing has occurred, in contrast to *S. latifolia*, which has a largely euchromatic Y chromosome (Grabowska-Joachimiak and Joachimiak 2002) or *R. hastatulus*, in which only the older Y chromosome is heterochromatic (Grabowska-Joachimiak et al. 2014). However, although this is an open question worth future study, lack of expression makes these genes at least functionally hemizygous, which is the crucial point for the questions addressed here.

Patterns of gene loss from the Y chromosome

Gene silencing and/or loss in *R. rothschildianus* is extensive and has occurred rapidly, but it is not rampant or random: genes that have been lost from the Y have lower ancestral expression and lower ancestral constraint than genes that remain on the Y (**table 4**). They are also less likely to be expressed in the haploid phase of the lifecycle (**table 5**), probably over and above any effect this has on the constraint of genes. Finally, genes that remain on the Y chromosome are enriched for various fundamental functional categories. The genes that still remain on the *R. rothschildianus* Y chromosome could reflect a core of genes that cannot be lost similar to that found in old, fully degenerated sex chromosomes, although the number of genes involved would be about 10 times higher than in mammals (~36 retained genes; Bellott et al. 2014) or *Drosophila* species (<20 retained genes; Bernardo Carvalho et al. 2009). This difference could be the result of more extensive haploid gene expression in plants increasing the number of genes that cannot be lost from the Y chromosome due to their being essential for the development of Y-bearing pollen.

However, the relative youth of sex chromosomes in *R. rothschildianus* (~8-10 million generations compared to, for example, ~160-180 million years ago in mammals; Potrzebowski et al. 2008; Veyrunes et al. 2008) suggests gene loss is probably on-going. This implies purifying selection on the Y chromosome has a greater role than just maintaining a core set of genes. Instead, purifying selection determines which genes are lost first, as opposed to it being a random mutation-driven process. A similar pattern has also been observed on the neo-sex chromosome of *D. miranda*, where genes that have been lost were ancestrally expressed at lower levels (Kaiser et al. 2011), and in threespine stickleback, where genes retained on the oldest stratum are under stronger purifying selection than the rest of the X chromosome (White et al. 2015). It is worth pointing out that these results also suggest that reports of the Faster-X effect, where hemizygous genes have a faster rate of evolution due to increased efficacy of positive selection

(Charlesworth et al. 1987), of partially degenerated sex chromosomes should be treated with caution when they come from comparisons between genes as opposed to between the same genes in different species. As well as sometimes being driven by neutral drift as opposed to positive selection (Wright et al. 2015), the Faster-X effect might also sometimes be driven by the inherently lower constraint in genes that are lost from the Y chromosome.

The pattern of gene loss observed in *R. rothschildianus* can be used to distinguish to some extent between possible models of Y chromosome degeneration. In particular, it does not support the idea that it is driven by active silencing of the Y allele due to either adaptive evolution of the X or mildly deleterious mutations on the Y, as these processes would be expected to first target genes undergoing extensive positive selection in the former or more important genes in the latter, which is the opposite pattern to the one found here. Additionally, although analysis of allele-specific expression suggests Y alleles are being expressed on average at a lower level than the X allele (**fig 8**), there is no clear pattern with age or degree of constraint, which suggests this might be a direct cause of deleterious mutations on the Y targeting regulatory regions, as opposed to any active suppression of particular genes. Arguably, adaptive silencing would also be more likely to be accompanied by extensive dosage compensation, as otherwise the fitness cost of a loss of a gene copy is likely to outweigh any benefit arising from suppression of a less fit allele. Instead, the pattern in types of genes that are lost or retained in *R. rothschildianus* suggests Y chromosome degeneration is a direct consequence of interference between sites reducing efficacy of selection. In particular, background selection could have a prominent role, as the genes that remain on the Y show evidence of being under purifying selection – they remain on the Y chromosome and they are more constrained than sex-linked XY genes in *R. hastatulus*. Purifying selection on these more constrained genes will have resulted in background selection reducing effective population size of non-recombining regions and increasing the speed of Muller's ratchet (Gordo and Charlesworth 2001), ultimately leading to the degeneration and loss of more weakly selected linked genes.

There is indirect evidence that sex-specific selection is resulting in adaptive evolution, suggesting gene loss might also be prevented by male-driven selection. Half of all genes that still retain a Y copy in *R. rothschildianus* have significantly different expression between males and females, considerably more than in autosomal genes (**table 6**). This is not surprising as sex chromosomes are likely to be hotspots for alleles that would otherwise be sexually antagonistic, and results from other taxa with partially degenerated sex chromosomes have found the same enrichment of differentially expressed genes (Vicoso, Kaiser, et al. 2013; White et al. 2015). Although other studies have reported a feminisation of the X (Sturgill et al. 2007, Allen et al. 2013), attributed to the fact that the X chromosome spends twice the amount of time in females as males, differences in gene expression in *R. rothschildianus* are biased towards a higher expression in males. This might be due to a time-dependent process, where initial masculinization is followed by feminization of the X chromosome (Zhou and Bachtrog 2012b).

Interestingly, this male bias in gene expression is arising from preferential expression of the Y allele (**fig 9**). These Y alleles did not necessarily arise on the Y chromosome, as they might have been present in the ancestral pseudo-autosomal region and therefore contributed to the suppression of recombination due to sexually antagonistic effects (Rice 1987a; Jordan and Charlesworth 2011). However, the highly skewed expression favouring the Y allele over the X allele must have occurred after recombination was suppressed. Although it is possible that mutations in regulatory regions is causing deleterious increases in expression level, as opposed to the more commonly reported reduction in expression, high expression levels of degenerating genes is likely to be strongly selected against. Moreover, the enrichment of these Y-biased genes for functions that are likely to have different optima between sexes, such as reproductive structure development and negative regulation of flower development, implies this is an adaptive bias in allele expression. Overall, this suggests that the Y chromosome is not only passively degenerating, but that it can also respond to male-specific selection. This also

provides a mechanism by which selective sweeps, the last of the non-interference models, could be operating.

Evolution of gene expression in response to gene loss

There is no evidence that complete chromosome-wide dosage compensation has evolved in *R. rothschildianus*, despite the extensive gene loss (**fig 4**). This means that a large number of genes are expressed at a lower level in males than the ancestral, presumably closer to optimal, expression. This is in accordance with reports of the lack of complete dosage compensation from a wide range of other species, including birds, many insects, fish and snakes (eg Ellegren et al. 2007; Uebbing et al. 2012; Vicoso, Emerson, et al. 2013; Chen et al. 2014; Mahajan and Bachtrog 2015; Walters et al. 2015). The fact that full dosage compensation fails to appear in many species implies either the fitness cost of a reduced expression level is not large or, alternatively, that dosage compensation is a difficult thing to achieve.

However, in *R. rothschildianus* many sex-linked hemizygous genes have expression levels in males that are greater than half the value of female expression, and tests for differential expression between males and females showed that there was no significant difference in ~25% of hemizygous genes (**fig 5, table 6**). This suggests targeted dosage compensation may be occurring for particular genes, as opposed to a chromosome-wide phenomenon. This is supported by the negative correlation between male/female ratio and the ancestral d_N/d_S ratio, as this implies there is targeted selection occurring on the expression levels of more constrained, and therefore presumably more important, genes. Because genes with lower d_N/d_S ratio have higher expression levels (**supplementary fig 2**) this is very unlikely to be an artefact of "automatic" dosage compensation, which might occur in genes with low expression levels (Malone et al. 2012). Additionally, there was no correlation between between-individual variance and degree of compensation, which means it is also unlikely to be an artefact of random noise. Similar to reports from the chicken sex chromosomes, where dosage compensation is occurring only in dosage sensitive

genes (Zimmer et al. 2016), these results suggest that active targeted dosage compensation is taking place in *R. rothschildianus* of particular important genes.

Genes that show no significant difference in expression between male and female expression in *R. rothschildianus* are often significantly below ancestral gene expression (**fig 7B**). However, they are invariably significantly above half the value of ancestral expression. This suggests that dosage compensation in *R. rothschildianus* might be occurring through down-regulation of the expression of one or both of the X chromosomes in females, as well as an up-regulation of the expression in males. This has been reported in other species that have either patchy or partial dosage compensation, such as Strepsistera beetles and *Heliconius* butterflies (Walters et al. 2007; Mahajan and Bachtrog 2015), as well as species with complete dosage compensation, such as mammals and *C. elegans* (Ercan et al. 2007; Pessia et al. 2012). Reduction of expression levels in the homogametic sex therefore appears to be a common, albeit counterintuitive, way of achieving within-genome network expression balance.

In mammals, it has been shown that autosomal genes that interact with these genes on the X chromosomes are down-regulated accordingly (Julien et al. 2012). Presumably, something similar is occurring in other taxa that exhibit this pattern. This suggests that achieving equal expression levels by simple up-regulation of a single copy of a gene is not trivial. Although the outwardly simplest mechanism of up-regulation of the hemizygous chromosome in the heterogametic sex has repeatedly evolved in Diptera (Vicoso and Bachtrog 2015), they are an anomaly in the range of widely varied mechanisms and degrees of dosage compensation in other taxa (Mank 2013). Moreover, even in Diptera genes with high levels of male-biased expression are under-represented on the X, which could result from a difficulty in achieving high expression levels from a single gene copy (Vicoso and Charleworth 2009). It may therefore be a difficulty in achieving high levels of gene expression from a single copy of a gene that results in the wide range of degrees and diverse set of approaches to dosage compensation observed in different taxa.

Methods

RNA sequencing and transcriptome assembly

Rumex rothschildianus is a rare dioecious annual endemic to Israel (Rottenberg and Parker 2003). Material of *R. rothschildianus* was obtained from a bulk seed collection from the Tel Aviv botanical garden. I collected leaf tissue from the parents, six female and six male offspring of two independent crosses of *R. rothschildianus*. These crosses were carried out in isolation in glasshouse conditions to prevent any possibility of cross-fertilization from an individual other than the focal male parent. The second cross only had four male offspring due to extreme sex ratio bias. I extracted total RNA using RNAeasy plant kit (Qiagen) following manufacturer recommended protocols. Illumina Hi-Seq 2500 sequencing of 100 bp paired-end reads was carried out at The Centre for Applied Genomics (Toronto, Canada) multiplexed across four lanes, with one lane having fewer samples allowing increased coverage of the two individuals (one male parent and one female parent) subsequently used for *de novo* transcriptome assembly.

I assembled separate female and male reference transcriptomes *de novo* using one female and one male parent. Prior to assembly the reads were passed through a quality filter to remove low quality read pairs with over 10% Ns, over 50% low quality bases, or shorter than 50bp long. The perl script VelvetOptimiser.pl (v. 2.2.4) available from GitHub was used to determine the optimal kmer size (45 for the female assembly; 27 for the male assembly) and the assembly was subsequently carried out using Velvet (v. 1.2.09; Zerbino and Birney 2008) followed by Oases (v. 0.2.08; Schulz et al. 2012). The output from Oases contains several isoforms per transcript, and in each case I chose the longest as the representative transcript. This pipeline resulted in a female reference transcriptome of 31408 contigs (N50=1894) and a male reference transcriptome of 58358 contigs (N50=1766). I obtained

reference transcriptomes for *R. hastatulus* and hermaphroditic *R. bucephalophorus* from Hough et al. (2014).

Identification of sex-linked genes

To identify genes located on the sex chromosomes in *R. rothschildianus* I mapped all individuals to the female reference transcriptome using Burrows-Wheeler Aligner (v. 0.7.8-r455; Li and Durbin 2009) followed by Stampy (v.1.0.20; Lunter and Goodson 2011) to map more divergent reads. I carried out a pre-processing pipeline on the mapped bam files using Picard tools (http://picard.sourceforge.net) as described in the GATK "best practices" (DePristo et al. 2011; Van der Auwera et al. 2013). I then used GATK haplotype caller followed by unified genotyper to call variants, implemented with the recommended quality filters (McKenna et al. 2010).

The resulting VCF files contained variant information for the parents and male and female offspring of two independent crosses. I parsed these files searching for SNPs segregating in diagnostic patterns to form three different datasets: genes located on the X and Y chromosomes (hereafter 'sex-linked XY genes'); genes located on the X and missing from the Y chromosome ('sex-linked hemizygous genes'); and genes located on the autosomes. Sex-linked XY genes can be identified by heterozygous sites in the father in which SNPs are inherited by all of the sons and none of the daughters (which must therefore be located on the Y chromosome) or in which they are inherited by all of the daughters and none of the sons (which must therefore be located on the X chromosome). Sex-linked hemizygous genes can be identified by sites that are homozygous in both the mother and the father, heterozygous in all of the daughters and homozygous for the mother's variant in all of the sons. I used this segregation pattern to get a "strict" set of hemizygous genes, which I used in the analyses of gene expression (see below). I also used a second less stringent hemizygous segregation pattern involving sites that are heterozygous in the mother, homozygous in the father, have both heterozygous and homozygous

daughters, and only have homozygous sons that exhibit both the female parent's alleles. This pattern is more susceptible to genotyping error, but because more than one diagnostic SNP is required to classify a gene into any particular class most genes identified using this pattern are still likely to be hemizygous. I identified autosomal genes using SNPs segregating in the usual Mendelian manner, in each case requiring at least two of the offspring of each sex to have inherited a SNP from both parents.

I classified a gene into a particular set (sex-linked XY, sex-linked hemizygous, autosomal) when it had at least 4 SNPs segregating in the diagnostic pattern and no SNP segregating in a conflicting pattern based on the first independent cross. I used the second cross to validate these genes by requiring at least one SNP segregating in the diagnostic manner for each gene and no SNPs segregating in a conflicting patter. I generated a "strict hemizygous" dataset for the analyses involving gene expression, where any one gene not being hemizygous could influence the conclusions, and where it was necessary to be equally confident of the hemizygosity in both independent crosses, as the analyses used expression data from all offspring. For this "strict hemizygous" dataset I required each gene to have 3 diagnostic SNPs in both crosses of only the more stringent segregation pattern (see above). I obtained lists of sex-linked XY genes, sex-linked hemizygous and autosomal genes for *R. hastatulus* from Hough et al. (2014).

Confirmation of hemizygosity

I confirmed hemizygosity of the genes identified as sex-linked hemizygous in *R. rothschildianus* in several ways to address two possible sources of error. The first of these is that the hemizygous segregation pattern could be result of high sequence divergence between the X and Y sequences resulting in a Y transcript failing to map to the female reference. To investigate this possibility I carried out a protein BLAST (Altschul et al. 1990) of the hemizygous genes against both the male and female reference transcriptome. I compared the top four BLAST hits between the two transcriptomes, looking for hits on the male reference that were not present in the

female reference, which would indicate closely related sequences that had assembled separately. I also used a second approach testing for this possibility, which involved mapping all individuals to the male reference transcriptome. I then identified male-specific genes by looking for transcripts that were present in all males but absent in all females. I defined these by requiring sites to be missing in all females for at least half of the length of the transcript. In addition, I removed any transcript with 4 or more SNPs between the father and the male offspring, as these would consist of genes uniquely expressed in males but not Y chromosome specific, as the variant must be inherited from the mother. This resulted in 99 male-specific transcripts, on which I then preformed a protein BLAST against the female reference and searched the resulting hits for evidence of similarity to genes in the hemizygous dataset. These two approaches lead to the removal of a total of four suspicious genes from the hemizygous dataset.

The second possible source of erroneous classification as hemizygous could arise from a transcript on the Y chromosome still being expressed, but at very low levels that are beneath the GATK cut-off for calling heterozygosity. To eliminate this possibility I conducted a manual examination of a random subset of these genes using the viewer IGV (Robinson et al. 2011). Additionally, I parsed allele-specific data from the VCF file to look for the presence of low levels of a second allele in hemizygous genes in males. However, there were no hints of heterozygosity in any hemizygous genes in males.

It should be noted that because all the data comes from transcriptomes it is not possible to distinguish genes that are truly removed from the Y chromosome as opposed to genes that are simply not being expressed on the Y chromosome. However, these confirmation steps provide confidence in the fact that genes are functionally hemizygous, with only one gene copy being expressed in males.

Identification of orthologs, construction of alignments and phylogenetic trees

To identify orthologs between *R. rothschildianus* and two closely related species *R. hastatulus* and *R. bucephalophorus* I first identified the longest open reading frame (ORF) of each transcript from each species using the program getorf from the software suite EMBOSS (v. 6.4.0.0; Rice et al. 2000). I then performed a reciprocal nucleotide BLAST search between each species pair, and used three-way best hits to determine orthology.

To place *R. rothschildianus* in the *Rumex* phylogeny I downloaded 10 *R. acetosa* genes and the transcriptome of *Fagopyrum esculentum*, in the same family as *Rumex* (Polygonaceae) and suitable for use an outgroup, from the GenBank nucleotide database. I identified the longest ORF for each sequence of both species and used reciprocal nucleotide blast searches of *R. acetosa* and reciprocal protein blast of *F. escalatum* against the longest ORFs of the three reference transcriptomes to determine orthology.

I constructed alignments of orthologous ORFs by first aligning the amino acid sequences using the program Muscle (v3.8.3; Edgar 2004), which were then used to guide the alignment of nucleotide sequences using the program RevTrans (v. 1.4; Wernersson and Pedersen 2003). I constructed maximum likelihood phylogenetic trees using the program RAxML (v. 8.2.4; Stamatakis 2014) and bootstrapped the trees to get measures of support for each node using the same program. The trees were rooted using *F. escalatum* as an outgroup.

Molecular evolution on the X and Y chromosomes

To investigate molecular evolution of the X and Y sequences it is necessary to have phased consensus sequences. This was carried out using a likelihood approach developed and tested in Hough et al (2014). Briefly, this involves an R script that assesses sex-linkage of SNPs using a likelihood ratio, and generates a consensus Y

sequence given the female reference as a consensus X sequence and the VCF file of variants mapped to it. I modified this script slightly to include the evolution of indels, but other details remain the same and are described in full in Hough et al. (2014). Because of the slight modification of the script I generated consensus phased X and Y sequences for both *R. rothschildianus* and *R. hastatulus.*

I identified the longest ORF separately for the X and Y consensus sequences and constructed pairwise alignments of the X and Y sequences following the same method described above. I calculated pairwise X-Y sequence divergence using the yn00 program implemented in PAML (v. 4.8; Yang 2007). The number of genes that were sex-linked in both *R. rothschildianus* and *R. hastatulus* was low (see below), so for each species I aligned the phased XY sequences for *R. rothschildianus* and *R. hastatulus* to their orthologs in the other species and the hermaphrodite *R. bucephalophorus*, and I analyzed these four-way alignments separately using the codeml program in PAML. Input tree files were unrooted, and in each case the X and Y sequences formed a monophyletic group and there was a trifurcation at the root. I used an initial run of the M0 "one-ratio" model to provide the initial branch lengths for the trees used in the subsequent analyses. I ran several models: 1) Branch models allowing the rate of the X or the Y sequence to vary compared to the "background" rate of the rest of the tree; 2) A free ratio model allowing rates to vary freely for each branch; 3) Branch-site models allowing rates to vary on particular branches as well as between sites, which is used to detect signals of positive selection when particular sites (codons) have d_N/d_S ratio greater than one.

I ran the same PAML models on sex-linked hemizygous genes and autosomal genes in *R. rothschildianus* and their orthologs in the other species. The analyses looking for evidence of positive selection included the out-group *F. escalatum*, so were also based on alignments of four sequences. I generated the tree file separately for each gene using the program RAxML, as the phylogenetic relationship between *R. rothschildianus*, *R. hastatulus* and *R. bucephalophorus* is not fully resolved.

Assessing inter-specific homology of sex chromosomes

To assess the possible homology of sex chromosomes in *R. rothschildianus* and *R. hastatulus* I carried out a thorough comparison of the all sex-linked genes in the two species using datasets of different stringency (using different number of SNPs required to define the gene as sex-linked) and different age (using X-Y synonymous divergence as a measure of age), looking for the presence of more orthologous genes present in each dataset than one would expect by chance. I calculated expected values simply by random expectation given the number of orthologs between the two species, and the number of orthologs in each gene class. I calculated significance of the overlap between datasets by two-tailed random permutation tests.

Analyses of gene and allele-specific expression

I measured levels of gene expression using the python package HTSeq (Anders et al. 2015), using the longest ORF as the defined region for expression counts. I analysed the output of raw counts using the R package DESeq2 (v. 1.11.45; Love et al. 2014), which uses a negative binomial distribution model to test for significantly different levels of gene expression between groups. I used this to identify genes with significantly different expression between 1) male and female individuals in *R. rothschildianus*, and 2) present expression levels in *R. rothschildianus* and ancestral expression levels, estimated by using the average of male and female expression in *R. hastatulus*. Because hemizygous genes are expected to have lower levels of expression in males relative to females I carried out the between-individual normalization using only the autosomal genes, which have no reason to be biased overall in their expression between the genders. I then used the individual size factors obtained from the autosomal data to transform the raw counts of sex-linked genes into normalized counts. None of the hemizygous genes in *R. rothschildianus* are hemizygous in *R. hastatulus*, allowing *R. hastatulus* to be used for estimation of ancestral expression levels. The close correlation between expression in *R. rothschildianus* and *R. hastatulus* (r_s=0.60, $P<10^{-15}$) validates this approach.

I obtained allele-specific expression from allele read counts in the VCF files. SNPs with low coverage (<8 reads) were removed, as these are more likely to give highly skewed ratios. I calculated the ratio of alternate over reference allele counts for each heterozygous individual: in males this corresponds to Y over X expression, whereas in females this corresponds to expression of the X inherited from the father over the X inherited from the mother. I then averaged these ratios across individuals and across SNPs in each gene, giving an average alternative/reference value per gene. I calculate the same alternate over reference allele count ratio for autosomal genes to provide the null expectation, as mapping bias lowers the actual expected ratio to slightly below the theoretical null expectation of 1:1.

Generating gametophyte and sporophyte expression gene lists

Lists of genes expressed during different life cycle stages were obtained for *Arabidopsis thaliana* and *Nicotiana tabacum* from Borges et al. (2008) and Hafidh et al. (2012). I divided these into two sets of genes: those expressed at any point during the haploid gametophyte phase of the life cycle, and those only expressed during the diploid sporophyte phase. I downloaded the sequences of these genes from GenBank, and carried out reciprocal protein blasts between *R. rothschildianus*, *A. thaliana* and *N. tabacum* to determine orthology. I generated a consensus list of genes expressed during the haploid phase of the lifecycle using the overlap of gene between *A. thaliana* and *N. tabacum*. I generated a list of genes expressed only in the sporophyte using the combined datasets from both species, as the lists were from expression profiles of different tissues (seedlings, leaves and roots) and therefore the overlap between them contained only a small number of genes. However, I removed any conflicting genes (genes expressed in the haploid phase of either species). I then used these consensus lists of genes in the analysis of *R. rothschildianus*. I analysed the number of orthologous sex-linked hemizygous genes and sex-linked XY genes present in each dataset (expressed or not expressed during

haploid phase of lifecycle) using a Fisher's exact test. I also carried out a second test between all sex-linked genes compared to all autosomal genes.

Functional enrichment analysis

I used the Database for Annotation, Visualization and Discovery (DAVID) (Huang et al. 2009a; Huang et al. 2009b) to identify significant overrepresentation of particular functional categories in genes still present on the Y chromosome compared to genes that have been lost from the Y chromosome. I also carried out another functional enrichment test on genes showing bias in expression towards the Y allele in males. Gene ontology was defined using orthology to *A. thaliana*, which has the most comprehensive gene annotation of any plant species. The background from which to base the enrichment analysis on was made up of all *R. rothschildianus-A. thaliana* orthologs. DAVID provides functional annotation clustering to facilitate interpretation, with enrichment scores above 1 being significant. Significance of enrichment of particular functional categories is determined using a modified (more conservative) Fisher's exact test.

References

Akagi T, Henry IM, Tao R, Comai L. 2014. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science*. 346:646–650.

Allen SL, Bonduriansky R, Chenoweth SF. 2013. The genomic distribution of sex-biased genes in *Drosophila serrata*: X chromosome demasculinization, feminization, and hyperexpression in both sexes. Genome Biol Evol. 5:1986–1994.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403-410.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106.

Anders S, Pyl PT, Huber W. 2015. HTSeq – a Python framework to work with high-throughput sequencing data. Bioinformatics. 21:166-169.

Bachtrog D. 2004. Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. Nat Genet. 36:518–522.

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. Nat Rev Genet. 14:113–124.

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice WR, Valenzuela N. 2011. Are all sex chromosomes created equal? Trends Genet. 27:350–357.

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T, Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature. 508:494–499.

Bergero R, Qiu S, Charlesworth D. 2015. Gene loss from a plant sex chromosome system. Curr Biol. 25:1234–1240.

Bernardo Carvalho A, Koerich LB, Clark AG. 2009. Origin and evolution of Y chromosomes: *Drosophila* tales. Trends Genet. 25:270–277.

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijó JA, Becker JD. 2008. Comparative transcriptomics of *Arabidopsis* sperm cells. Plant Physiol. 148:1168–1181.

Bull J. 1983. Evolution of sex determining mechanisms. Menlo Park (CA): Benjamin-Cummings Publishing Company.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B. 355:1563–1572.

Charleworth B, Charlesworth D. 2010. Elements of evolutionary genetics. Greenwood Village (CO): Roberts and Company Publishers.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 130:113–146.

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

Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, Song W, An N, Chalopin D, Volff J-N, et al. 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat Genet. 46:253–260.

Chibalina MV, Filatov DA. 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. Curr Biol. 21:1475–1479.

Comeron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity. 100:19–31.

Correns C. 1928. Bestimmung, Vererbung und Verteilung des Geschlechtes bei den hoheren Pflanzen. Vererbungswissenschaft. 2:1–138.

Cuñado N, Navajas-Perez R, de la Herran R, Ruiz Rejon C, Ruiz Rejon M, Santos J, Garrido-Ramos M. 2007. The evolution of sex chromosomes in the genus *Rumex* (Polygonaceae): identification of a new species with heteromorphic sex chromosomes. Chromosom Res. 15:825–832.

DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43:491-498

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res. 32:1792-1797.

Ellegren H, Hultin-Rosenberg L, Brunström B, Dencker L, Kultima K, Scholz B. 2007. Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biol. 5:40.

Ercan S, Giresi PG, Whittle CM, Zhang X, Green RD, Lieb JD. 2007. X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. Nat Genet. 39:403–408.

Field DL, Pickup M, Barrett SCH. 2012a. Comparative analyses of sex-ratio variation in dioecious flowering plants. Evolution. 67:661–672.

Field DL, Pickup M, Barrett SCH. 2012b. The influence of pollination intensity on fertilization success, progeny sex ratio, and fitness in a wind-pollinated, dioecious plant. Int J Plant Sci. 173:184–191.

Fisher R. 1930. The genetical theory of natural selection. Oxford: Clarendon Press.

Geraldes A, Hefer CA, Capron A, Kolosova N, Martinez-Nuñez F, Soolanayakanahally RY, Stanton B, Guy RD, Mansfield SD, Douglas CJ, et al. 2015. Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Mol Ecol. 24:3243–3256.

Gordo I, Charlesworth B. 2001. The speed of Muller's ratchet with background selection, and the degeneration of Y chromosomes. Genet Res. 78:149–161.

Grabowska-Joachimiak A, Joachimiak A. 2002. C-banded karyotypes of two *Silene* species with heteromorphic sex chromosomes. Genome. 252:243–252.

Grabowska-Joachimiak A, Kula A, Książczyk T, Chojnicka J, Sliwinska E, Joachimiak AJ. 2014. Chromosome landmarks and autosome-sex chromosome translocations in *Rumex hastatulus*, a plant with XX/XY1Y2 sex chromosome system. Chromosom Res. 23:187–97.

Hafidh S, Breznenová K, Růžička P, Feciková J, Čapková V, Honys D. 2012. Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. BMC Plant Biol. 12:24.

Hill W, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8:269–294.

Honys D, Twell D. 2004. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biol. 5:R85.

Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. 2014. Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. Proc Natl Acad Sci U S A. 111:7713–7718.

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

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37:1-13.

Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 4:44-57.

Jordan CY, Charlesworth D. 2011. The potential for sexually antagonistic polymorphism in different genome regions. Evolution. 66:505–516.

Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, Daish T, Grützner F, Kaessmann H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. PLoS Biol. 10:e1001328.

Kaiser VB, Charlesworth B. 2010. Muller's ratchet and the degeneration of the *Drosophila miranda* neo-Y chromosome. Genetics. 185:339–348.

Kaiser VB, Zhou Q, Bachtrog D. 2011. Nonrandom gene loss from the *Drosophila miranda* neo-Y chromosome. Genome Biol Evol. 3:1329–1337.

Kejnovský E, Michalovova M, Steflova P, Kejnovska I, Manzano S, Hobza R, Kubat Z, Kovarik J, Jamilena M, Vyskot B. 2013. Expansion of microsatellites on evolutionary young Y chromosome. PLoS One. 8:e45519.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 25:1754-1760.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 15:550.

Lunter G, Goodson M. 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21:936-939.

Mahajan S, Bachtrog D. 2015. Partial dosage compensation in Strepsiptera, a sister group of beetles. Genome Biol Evol. 7:591–600.

Malone JH, Cho D-Y, Mattiuzzo NR, Artieri CG, Jiang L, Dale RK, Smith HE, McDaniel J, Munro S, Salit M, et al. 2012. Mediation of Drosophila autosomal dosage effects and compensation by network interactions. Genome Biol. 13:R28.

Mank JE. 2013. Sex chromosome dosage compensation: definitely not for everyone. Trends Genet. 29:677–683.

Matsubara K, Tarui H, Toriba M, Yamada K, Nishida-Umehara C, Agata K, Matsuda Y. 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. Proc Natl Aca Sci U S A. 103:18190–18195.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297-1303

Melamed E, Arnold AP. 2007. Regional differences in dosage compensation on the chicken Z chromosome. Genome Biol. 8:R202

Ming R, Bendahmane A, Renner SS. 2011. Sex chromosomes in land plants. Annu Rev Plant Biol. 62:485–514.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 1:2–9.

Navajas-Perez R, de la Herran R, Lopez Gonzalez G, Jamilena M, Lozano R, Ruiz Rejon C, Ruiz Rejon M, Garrido-Ramos MA. 2005. The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastidial sequence data. Mol Biol Evol. 22:1929–1939.

Ogawa A, Murata K, Mizuno S. 1998. The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. Proc Natl Acad Sci U S A. 95:4415–4418.

Orr HA, Kim Y. 1998. An Adaptive Hypothesis for the Evolution of the Y Chromosome. Genetics. 150:1693–1698.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science. 327:92–94.

Papadopulos AST, Chester M, Ridout K, Filatov DA. 2015. Rapid Y degeneration and dosage compensation in plant sex chromosomes. Proc Natl Acad Sci U S A. 112:13021–13026.

Parker JS, Clark MS. 1991. Dosage sex-chromosome systems in plants. Plant Sci. 80:79–92.

Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GAB. 2012. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. Proc Natl Acad Sci U S A. 109:5346–5351.

Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jégou B, Kaessmann H. 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biol. 6:709–716.

Quesada del Bosque M, Navajas-Perez R, Panero J, Fernandez-Gonzalez A, Garrido-Ramos M. 2011. A satellite DNA evolutionary analysis in the North American endemic dioecious plant *Rumex hastatulus* (Polygonaceae). Genome. 54:253–260.

Rice WR. 1987a. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution. 41:911–914.

Rice WR. 1987b. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics. 116:161–167.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16: 276-277.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2012. Integrative Genomics Viewer. Nat Biotechnol. 29:24-26

Rottenberg A, Parker JS. 2003. Conservation of the critically endangered *Rumex rothschildianus* as implied from AFLP diversity. Biol Conserv. 114:299-303

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 28:1086-1092

Singh ND, Koerich LB, Carvalho AB, Clark AG. 2014. Positive and purifying selection on the *Drosophila* Y chromosome. Mol Biol Evol. 31:2612–2623.

Smith BW. 1969. Evolution of sex-determining mechanisms in *Rumex*. Chromosom Today. 2:172–182.

Stamatakis A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics. 30:1312–1313.

Steflova P, Hobza R, Vyskot B, Kejnovsky E. 2013. Strong Accumulation of Chloroplast DNA in the Y Chromosomes of *Rumex acetosa* and *Silene latifolia*. Cytogenet Genome Res. 141:59-65.

Steflova P, Tokan V, Vogel I, Lexa M, Macas J, Novak P, Hobza R, Vyskot B, Kejnovsky E. 2013. Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY1Y2) in the dioecious plant *Rumex acetosa*. Genome Biol Evol. 5:769–782.

Stehlik I, Barrett SCH. 2006. Pollination intensity influences sex ratios in dioecious *Rumex nivalis*, a wind-pollinated plant. Evolution. 60:1207–1214.

Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the *Drosophila* genus. Nature. 450:238–241.

Van der Auwera GA, Carneiro M, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J. et al. 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinformatics. 43:1-33.

VanBuren R, Zeng F, Chen C, Zhang J, Wai CM, Han J, Aryal R, Gschwend AR, Wang J, Na J, et al. 2015. Origin and domestication of papaya Y^h chromosome. Genome Res. 25:524–533.

Veyrunes F, Waters PD, Miethke P, Murchison EP, Kheradpour P, Sachidanandam R, Park J, Semyonov J, Chang CL, Whittington CM, et al. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Res. 18:965–973.

Vicoso B, Bachtrog D. 2009. Progress and prospects toward our understanding of the evolution of dosage compensation. Chromosom Res. 17:585–602.

Vicoso B, Bachtrog D. 2015. Numerous Transitions of Sex Chromosomes in Diptera. PLoS Biol. 13:e1002078.

Vicoso B, Charlesworth B. 2009. The deficit of male-biased genes on the *D. melanogaster* X chromosome is expression-dependent: a consequence of dosage compensation? J Mol Evol. 68:576–583

Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11:e1001643.

Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc Natl Acad Sci U S A. 110:6453–6458.

Walters JR, Hardcastle TJ, Jiggins CD. 2015. Sex chromosome dosage compensation in *Heliconius* butterflies: global yet still incomplete? Genome Biol Evol. 7:2545–2559.

Wernersson R, Pedersen AG. 2003. RevTrans - Constructing alignments of coding DNA from aligned amino acid sequences. Nucl Acids Res. 31:3537-3539.

White MA, Kitano J, Peichel CL. 2015. Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. Mol Biol Evol. 32:1981–1995.

Wilby AS, Parker JS. 1988. Recurrent patterns of chromosome variation in a species group. 61:55–62.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586-1591.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 12:32-43.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829

Zhou Q, Bachtrog D. 2012a. Chromosome-wide gene silencing initiates Y degeneration in *Drosophila*. Curr Biol. 22:522–525.

Zhou Q, Bachtrog D. 2012b. Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in *Drosophila*. Science. 337:341–346.

Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MTP, Zhang G. 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science. 346:1246338.

Zimmer F, Harrison PW, Dessimoz C, Mank JE. 2016. Compensation of dosage-sensitive genes on the chicken Z chromosome. Genome Biol Evol. 8:1233-2142

Figures and tables

Figure 1. Pairwise synonymous site divergence between X and Y sequences in sex-linked genes that still retain expression of the Y copy in *Rumex rothschildianus* and *R. hastatulus.*

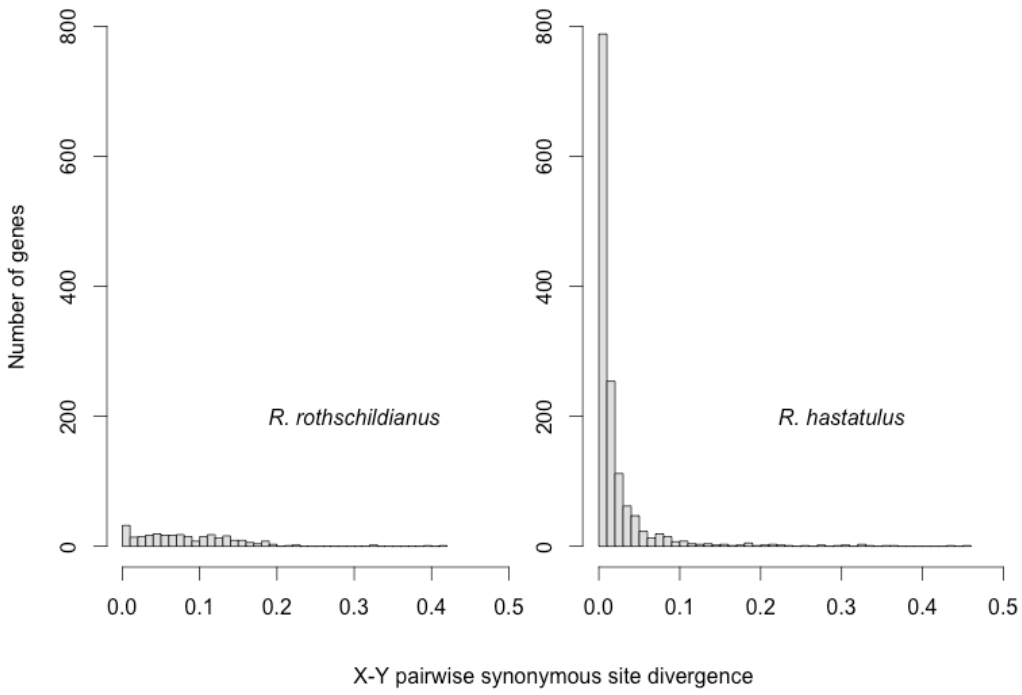

Figure 2. Pairwise non-synonymous site divergence versus synonymous site divergence between X and Y sequences in sex-linked genes that still retain expression of the Y copy in *Rumex rothschildianus* and *R. hastatulus.*

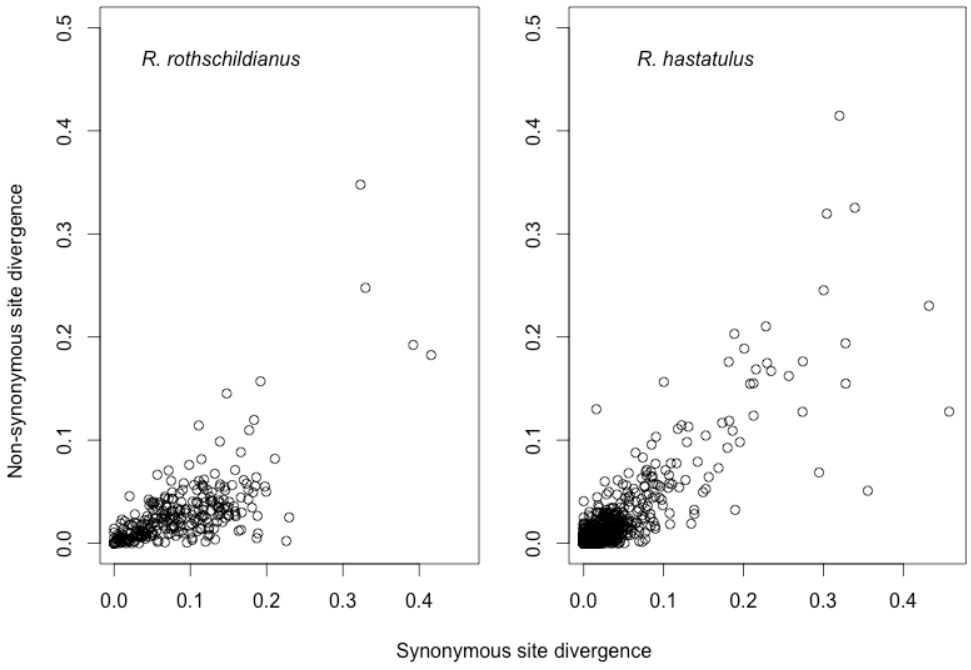

Figure 3. Comparison of gene expression in difference sets of genes. Male expression over female expression in *Rumex rothschildianus* and in *R. hastatulus* (the latter species used as a measure of ancestral expression) for genes that in *R. rothschildianus* are hemizygous (A) and autosomal (B). *R. rothschildianus* expression over ancestral expression for males and females for genes that in *R. rothschildianus* are hemizygous (C) and autosomal (D). Lines show 1:1 line expected if expression levels are equal between the two sets of genes compared and 0.5:1 line expected if expression level in the first set is half the value of second set.

.

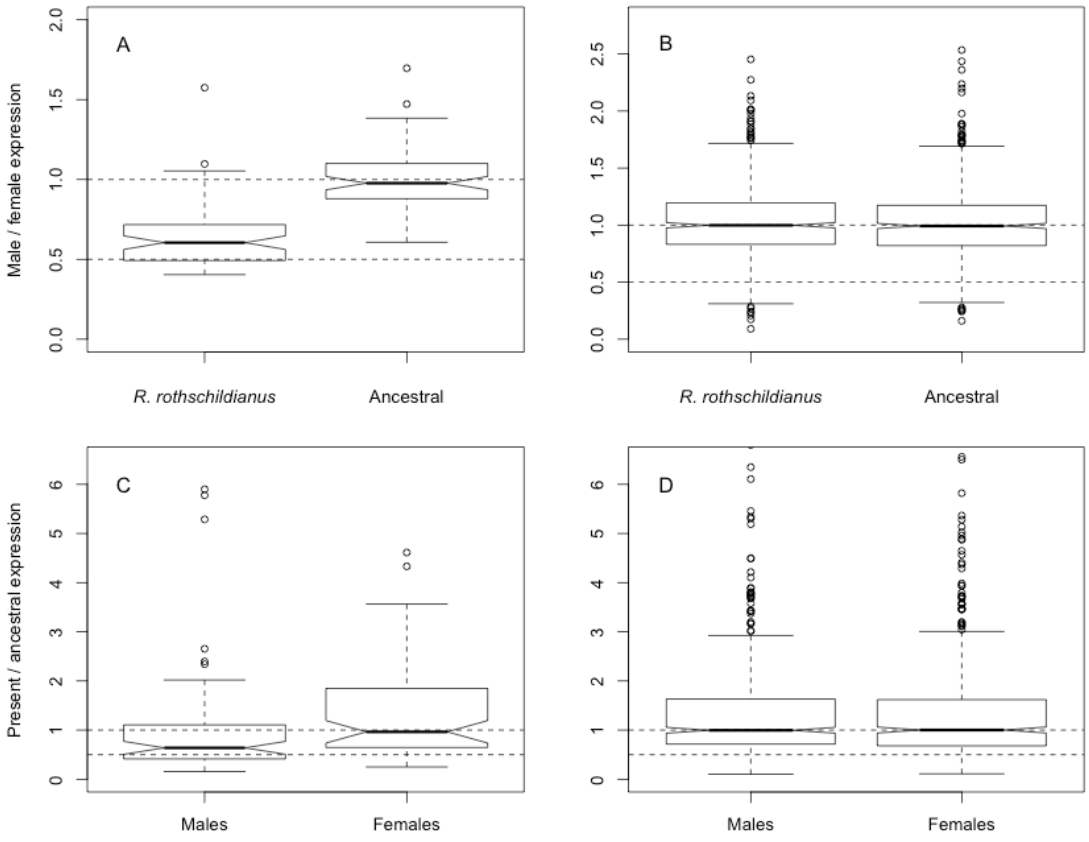

Figure 4. Male versus female expression level for hemizygous genes in *R. rothschildianus*. Lines showing 1:1 expression levels and 0.5:1 expression levels expected under scenarios of full dosage compensation (equal expression in both sexes) or no dosage compensation (half the expression level in males compared to females).

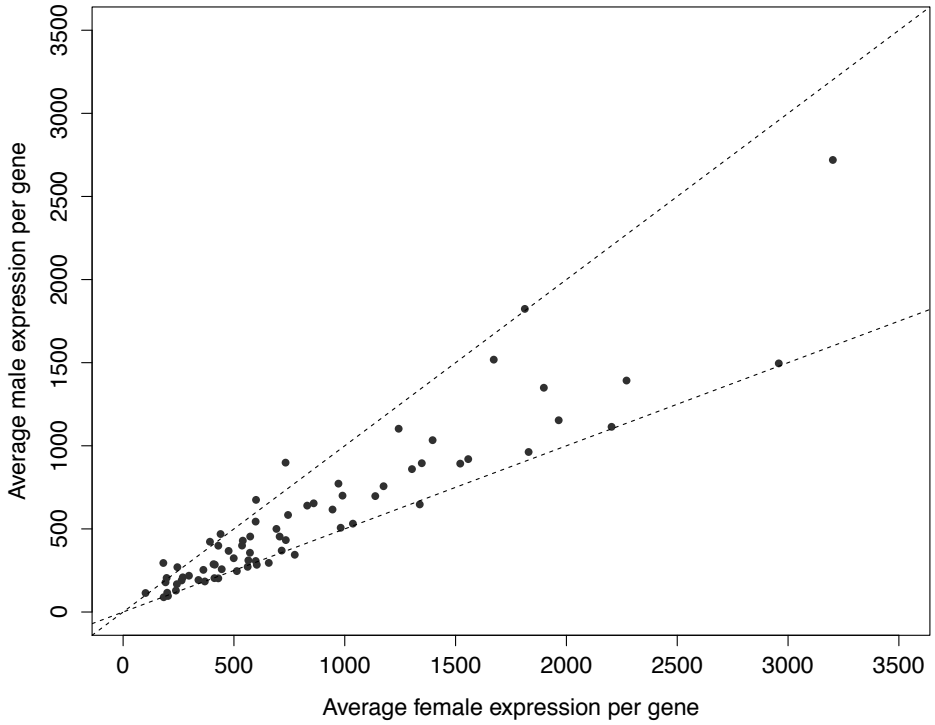

Figure 5. Significant negative correlation between the ratio of male expression over female expression and the d_N/d_S ratio in hemizygous genes, suggesting more constrained genes have been dosage compensated. Lines show 1:1 and 0.5:1 expression expected under scenarios of full dosage compensation or no dosage compensation.

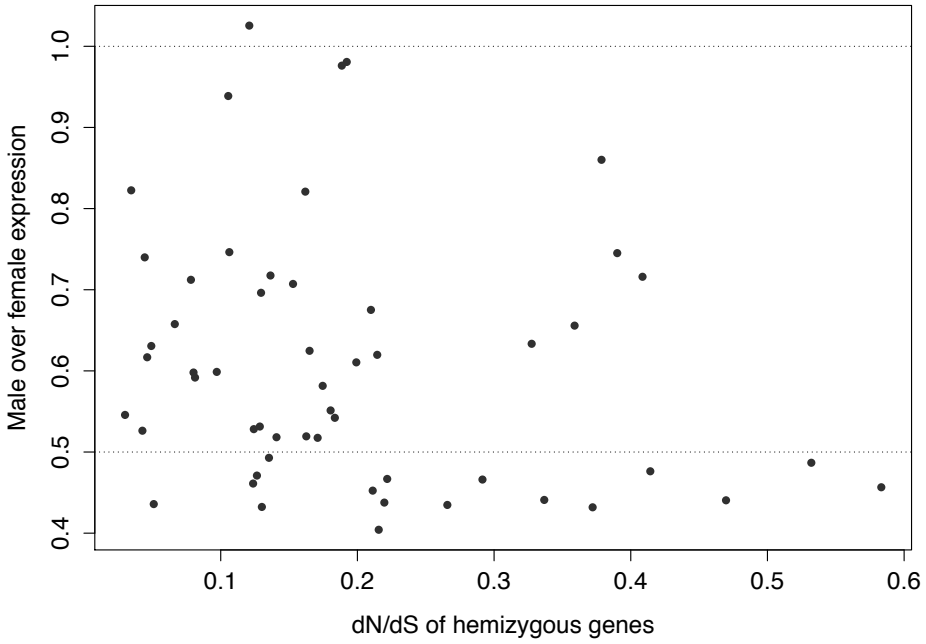

Figure 6. Female and male expression levels in *Rumex rothschildianus* over ancestral expression levels (given by *R. hastatulus*) for genes classified as "not compensated" (A) and "compensated" (B) based on the differential expression analysis to test for significantly different expression levels between males and females. Lines show 1:1 expression expected if expression has not changed from ancestral levels and 0.5:1 if current expression is equal to half the ancestral expression levels.

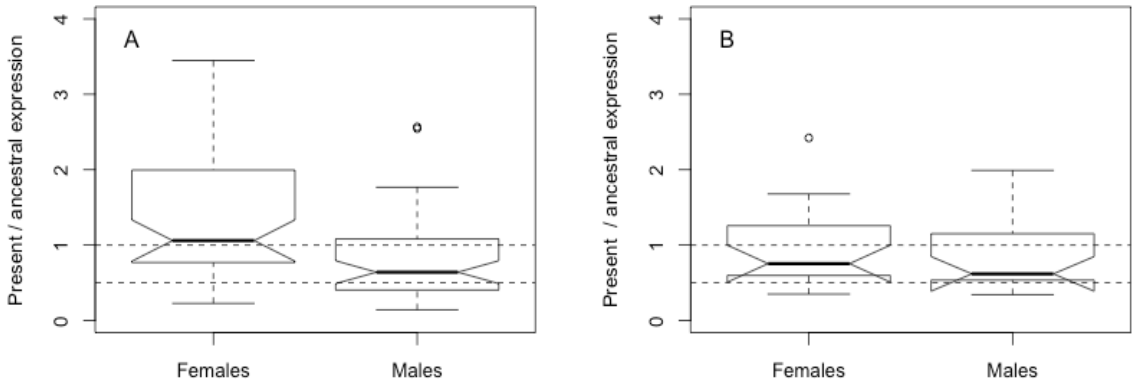

Figure 7. Allele-specific expression for sex-linked genes that still retain the Y copy and for autosomal genes in *Rumex rothschildianus*. Ratio of alternate/reference is shown for males (corresponding to the ratio of Y allele over X allele), females and autosomal genes in both males and females as a control. Line represents theoretical expectation of equal expression of both alleles, although mapping bias can shift the expected ratio to slightly below one.

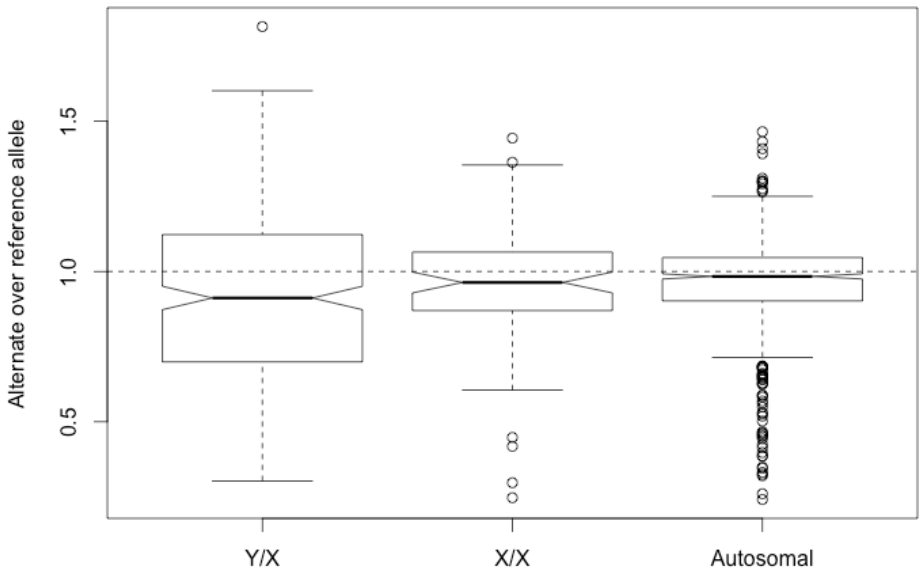

Figure 8. Ratio of male gene expression over female gene expression plotted against Y/X allele-specific expression ratio in *Rumex rothschildianus*, showing that higher expression levels in males is driven by over-expression of the Y allele. Lines represent neutral expectation of 1:1 expression between males and females and between X and Y alleles.

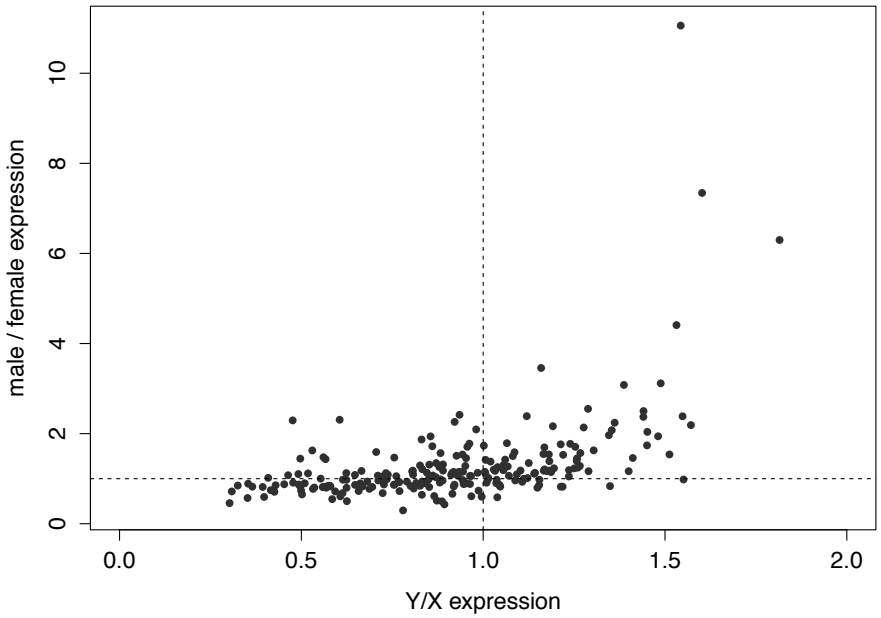

Table 1. Number of genes identified in this study as sex-linked XY (still present on the Y chromosome) and sex-linked hemizygous (lost from the Y chromosome) in *Rumex rothschildianus* and *R. hastatulus*. The estimated percentage gene loss is based on the number of sex-linked XY genes identified using SNPs that constituted polymorphisms on the X as opposed SNPs that were differences between the X and the Y.

	R. rothschildianus	***R. hastatulus***
Sex-linked XY	304	1545
Sex-linked hemizygous	249	142
Estimated gene loss	92%	18%

Table 2. Number of genes found in the same gene set in both *Rumex rothschildianus* and *R. hastatulus* and the number expected to overlap by chance. Old sex-linked genes are defined as those with X-Y synonymous divergence d_S>0.1.

	Expected	**Observed**
All sex-linked	40.7	40
Old sex-linked XY and hemizygous	0.165	0
Sex-linked (*R. roth*) vs autosomal (*R. hast*)	37.8	83
Sex-linked (*R. hast*) vs autosomal (*R. roth*)	90.6	98

Note. *R. roth, R. rothschildianus. R. hast, R. hastatulus.*

Table 3. Molecular evolution on the X and Y chromosomes in *Rumex rothschildianus* and *R. hastatulus.* PAML branch models were used to determine whether rate of X or Y varied significantly from the background rate. The d_N/d_s values were estimated from the 'free ratios' model in PAML. Genes with very low synonymous divergence (d_S<0.001) were removed for calculation of the mean as they give highly skewed estimates of d_N/d_s

	R. rothschildianus		***R. hastatulus***	
	X	**Y**	**X**	**Y**
Number passing BM LRT (%)	33 (16%)	71 (35%)	56 (5%)	172 (16%)
% faster than background	21%	94%	67%	93%
mean (SE) d_N/d_s	0.23 (0.02)	0.39 (0.02)	0.22 (0.02)	0.48 (0.03)
median d_N/d_s	0.15	0.31	0.09	0.32

Note. BM LRT, branch model likelihood ratio test. SE, standard error.

Table 4. Molecular evolution of different genes sets in *Rumex rothschildianus* and their orthologs in *R. hastatulus* and *R. bucephalophorus.* All values are mean d_N/d_S and associated standard error in brackets. Sex-linked XY genes are still present on the Y chromosome whereas sex-linked hemizygous have been lost from the Y chromosome.

Gene set in *R. rothschildianus*	*R. rothschildianus*	*R. hastatulus*	*R. bucephalophorus*
Hemizygous	0.225 (0.012)	0.236 (0.012)	0.231 (0.012)
Sex-linked XY	0.201 (0.010)	0.206 (0.010)	0.197 (0.010)
Autosomal	0.155 (0.006)	0.208 (0.006)	0.196 (0.006)

Table 5. Number of genes in different gene sets in *Rumex rothschildianus* that are expressed or not during the haploid phase of the lifecycle, based on data from *Arabidopsis thaliana* and *Nicotiana tabacum*. Sex-linked XY genes are still present on the Y chromosome whereas sex-linked hemizygous have been lost from the Y chromosome.

	Haploid expressed	**Not haploid expressed**
Sex-linked XY	60	70
Sex-linked hemizygous	29	83
Autosomal	129	240

Table 6. Number of genes that have significantly different expression between males and females in *R. rothschildianus* in different gene sets. Sex-linked XY genes are still present on the Y chromosome whereas sex-linked hemizygous have been lost from the Y chromosome. Significance was calculated using differential expression analysis based on a negative binomial distribution model

	Hemizygous	**Sex-linked**	**Autosomal**
% significantly different	76%	50%	14%
Number lower in males / total	52/53	49/125	36/80
% significantly different from 1/2 female expression	29%		
Number lower in males from 1/2 female expression / total	0/22		

Supplementary information

Figure S1. Significant negative correlation between pairwise d_N/d_S between X and Y sequences and the ancestral levels of gene expression in *Rumex rothschildianus* (r_S= -0.22, P<0.001) and *R. hastatulus* (r_S= -0.17, P<10^{-6}).

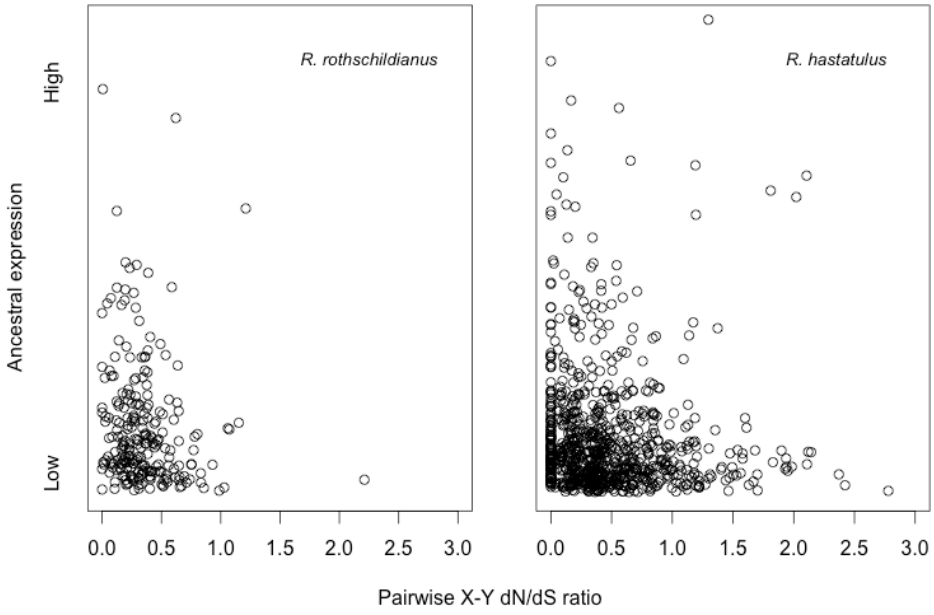

Figure S2. Significant negative correlation between d_N/d_S of hemizygous genes and female expression level (r_S= -0.32, *P*<0.02) in *Rumex rothschildianus*

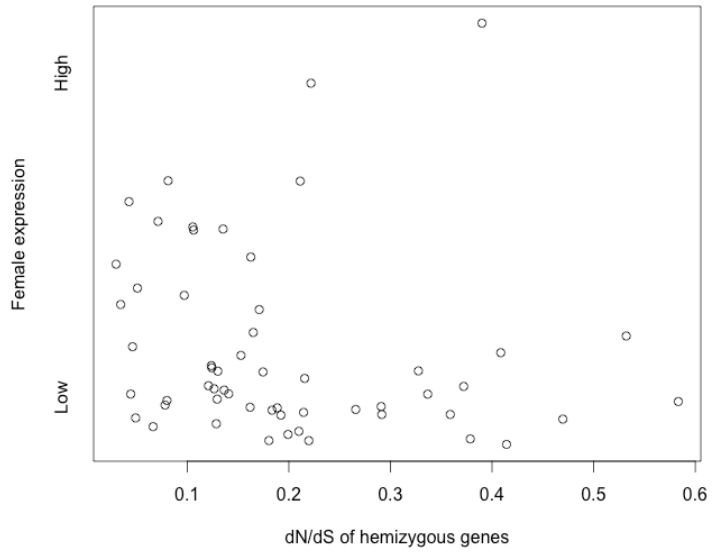

Figure S3. Male expression over female expression ratio is not significantly correlated with female expression (r_S= -0.020, *P*=NS) in *Rumex rothschildianus*

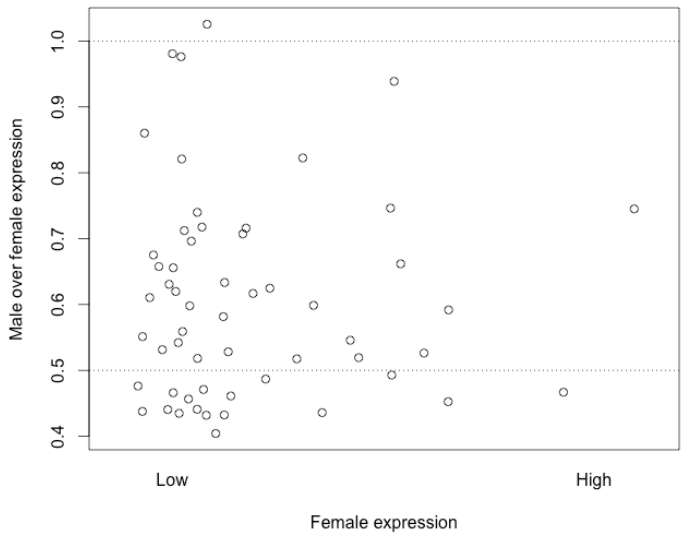

Figure S4. Male versus female expression ratio per gene for sex-linked XY genes (that still retain expression of the Y copy) in *Rumex rothschildianus.* Lines show 1:1, 0.5:1 and 2:1 expression ratios.

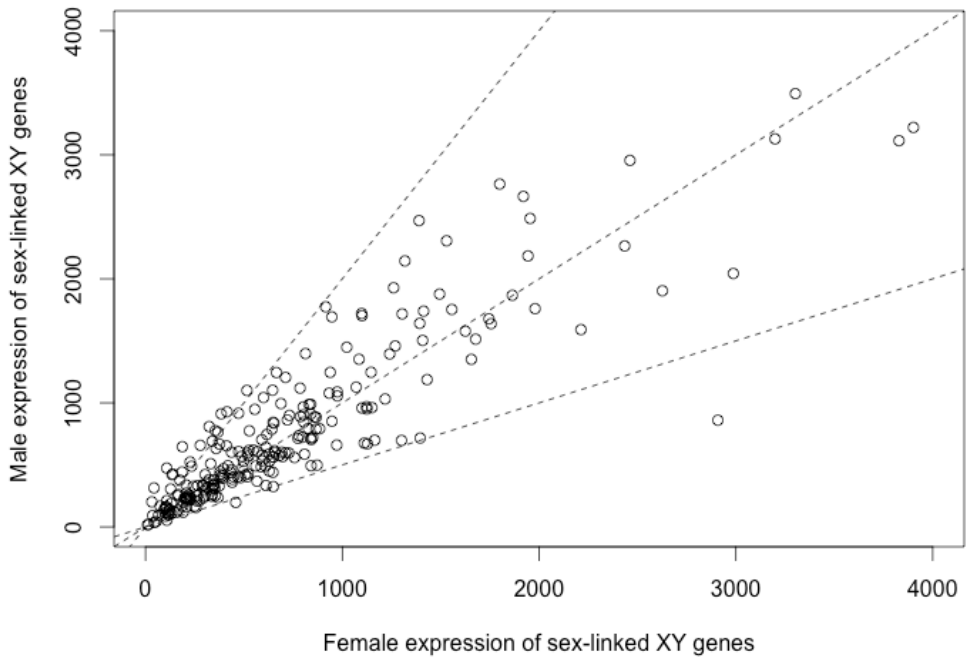

Table S1. Data for the two separate crosses in *Rumex rothschildianus* and the number of genes given by requiring different number of SNPs as a cutoff for classifying them into a particular gene set.

	Sex-linked XY			Sex-linked hemizygous			Autosomal		
Number of SNPs	**Cross 1**	**Cross 2**	**Shared**	**Cross 1**	**Cross 2**	**Shared**	**Cross 1**	**Cross 2**	**Shared**
1	526	497	391	1031	915	600	2027	3120	1373
2	451	418	352	559	492	389	1251	2047	798
3	405	383	321	380	361	288	883	1496	534
4	368	360	302	282	274	219	674	1170	389

Table S2. Data for overlap between sex-linked genes in *Rumex rothschildianus* and *R. hastatulus* using different numbers of SNPs as cut-off values for defining a gene as sex-linked. In *R. rothschildianus* the particular number of SNPs is required in both independent crosses.

Number of SNPs	***R. rothschildianus***	***R. rothschildianus* with orthologs**	***R. hastatulus***	***R. hastatulus* with orthologs**	**Number overlapping**	**Number expected by chance**
1	952	696	4512	2656	130	143.4
2	719	568	3007	1808	77	79.7
3	598	482	2178	1358	48	50.8
4	517	426	1703	1095	29	36.2

Table S3.
To place *R. rothschildianus* in the *Rumex* phylogeny we downloaded 10 sequences from *R. acetosa* and the *Fagopyrum esculentum* transcriptome. There were 5 genes that gave high (>80) bootstrap support of all nodes. Of these:
3 trees were (((ROTH,ACETOSA),(BUCE,HAST)),FAG);
1 tree was ((((ROTH,ACETOSA),HAST),BUCE),FAG);
1 tree was (((BUCE,HAST),ROTH),ACETOSA),FAG);

To determine whether the hermaphrodite *R. bucephalophorus* is an outgroup to the two clades with sex chromosomes we constructed trees using *F. esculentum* as an outgroup. There were 1368 genes that gave high (>80) bootstrap support of all nodes. Of these:
661 trees were (((BUCE,HAST),ROTH),FAG);
386 trees were (((BUCE,ROTH),HAST),FAG);
321 trees were (((HAST,ROTH),BUCE),FAG);

Tables S4 S5 and S6 were attached as separated excel spreadsheet.