The large-X effect in plants: Increased species divergence and reduced gene flow on the *Silene* X-chromosome

Xin-Sheng Hu, Dmitry A Filatov*

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX13RB, United Kingdom

* Corresponding author:

E-mail: dmitry.filatov@plants.ox.ac.uk

Phone: (44)1865 275 051

Fax: (44)1865 275 074

Running title: Large-X effect in dioecious Silene

Key words: speciation, sex chromosomes, large-X effect, Silene, gene flow

Abstract:

The disproportionately large involvement of the X-chromosome in the isolation of closely related species (the large-X effect) has been reported for many animals, where X-linked genes are mostly hemizygous in the heterogametic sex. The expression of deleterious recessive mutations is thought to drive the frequent involvement of the X-chromosome in hybrid sterility, as well as to reduce interspecific gene flow for X-linked genes. Here, we evaluate the role of the X-chromosome in the speciation of two closely related plant species – the white and red campions (*Silene latifolia* and *S. dioica*) – that hybridize widely across Europe. The two species evolved separate sexes and sex chromosomes relatively recently (~10^7 years), and unlike most animal species, most X-linked genes have intact Y-linked homologs. We demonstrate that the X-linked genes show a very small and insignificant amount of interspecific gene flow, while gene flow involving autosomal loci is significant and sufficient to homogenise the gene pools of the two species. These findings are consistent with the hypothesis of the large-X effect in *Silene* and comprise the first report of this effect in plants. Non-hemizygosity of many X-linked genes in *Silene* males indicates that exposure of recessive mutations to selection may not be essential for the occurrence of the large-X effect. Several possible causes of the large-X effect in *Silene* are discussed.

Introduction

The importance of sex chromosomes in speciation is reflected in the "two rules of speciation" (Coyne & Orr 1989): Haldane's rule and the large-X effect. Haldane's rule (Haldane 1922) states that the inviability or infertility of inter-specific hybrids is predominantly observed in the heterogametic sex. The large-X effect (also called the "Coyne rule", Turelli & Moyle 2007) concerns the reduced interspecific gene flow for X(or Z)-linked genes compared with that for autosomal genes (Dod *et al.* 1993; Ellegren & Parsch 2007; Ellegren *et al.* 2012; Macholan *et al.* 2007; Payseur *et al.* 2004; Teeter *et al.* 2008; Tucker *et al.* 1992), and the frequent involvement of the X-chromosome in hybrid sterility in animals (Coyne & Orr 2004; Jiggins *et al.* 2001; Masly & Presgraves 2007; Presgraves & Orr 1998). A consensus from the literature is that Haldane's rule is reasonably well explained by partial recessivity of hybrid incompatibility alleles that exhibit their full effects only in the heterogametic sex, the so called dominance theory (Coyne & Orr 2004; Turelli & Orr 1995). Causes of the large-X effect remain unclear, but a detailed genetic dissection of the two speciation rules in *Drosophila* indicated that the large-X effect could arise from a higher density of hybrid incompatibility loci on the X-chromosome (Masly & Presgraves 2007). Other possible reasons include partial recessivity of genic incompatibilities that are responsible for postzygotic isolation (Turelli & Orr 2000), meiotic drive of the sex chromosomes (Frank 1991; Hurst & Pomiankowski 1991; Tao & Hartl 2003), X chromosome mis-regulation in hybrids (Masly & Presgraves 2007), and faster-X evolution that is predicted to occur if adaptive mutations are partially recessive (Charlesworth *et al.* 1987; Vicoso & Charlesworth 2006).

Both Haldane's rule and the large-X effect have been reported in many animal species (Coyne & Orr 1989), but not in plants, with one notable exception. Haldane's rule was reported to occur in a pair of dioecious *Silene* species (Brothers & Delph 2010; Demuth *et al.* 2014), which evolved separate sexes and sex chromosomes only 10 million years ago (Charlesworth 2008), or even more recently (Rautenberg *et al.* 2012). However, the genetic

architecture underlying Haldane's rule in *Silene* appears to differ from those commonly found in animals (Demuth *et al.* 2014). This may partly be due to the fact that, unlike most animals, the *Silene* X-chromosome is largely non-hemizygous in males, as most X-linked genes still have functional Y-linked homologs (Bergero & Charlesworth 2011; Chibalina & Filatov 2011; Muyle *et al.* 2012). Thus, in *Silene,* partially recessive alleles of hybrid incompatibility are not expected to exhibit their full effects in the heterogametic sex (males), contradicting the dominance theory of Haldane's rule. Explanations for the existence of the "two speciation rules" have been primarily driven by studies in animals. Investigations of these rules in plants can provide important insights into the generality of these theories.

Here we examined the second rule of speciation by using two closely related dioecious plant species, *Silene latifolia* Poiret (white campion) and *Silene dioica* (L.) Clairv. (red campion). *S. latifolia* is a model species for research in various fields (Bernasconi *et al.* 2009), including the genetics of sex determination (Grant *et al.* 1994; Shull 1910; van Nigtevecht 1966a, b), evolution of sex chromosomes (Bergero *et al.* 2013; Chibalina & Filatov 2011; Qiu *et al.* 2013; Westergaard 1958), host–pathogen interactions (de Vienne *et al.* 2009), quantitative genetics (Brothers & Delph 2010; Demuth *et al.* 2014; Scotti & Delph 2006), biogeography (Prentice *et al.* 2008), evolutionary genetics and the ecology of adaptation (Muir *et al.* 2011; Muir & Filatov 2007) and speciation (Baker 1948, 1951; Minder *et al.* 2007; Muir *et al.* 2012; Waelti *et al.* 2008). *S. latifolia* and *S. dioica* are common across Europe, and hybridization between sympatric and parapatric natural populations has been observed (Baker 1948; Minder *et al.* 2007). The progeny of crosses between these species are healthy and fertile, though quantitative analyses reveal some reduction in pollen viability in hybrids compared to intra-specific crosses (Brothers & Delph 2010). *S. latifolia* inhabits open sunny fields and meadows, often growing along roadsides and in disturbed habitats. *S. dioica* prefers shady forests and moist soil. Nevertheless, the two species often come in contact and form hybrid swarms, resulting in significant inter-specific gene flow in contact areas (Minder *et al.* 2007). Despite this inter-specific gene exchange, *S. latifolia* and *S.*

dioica are clearly distinguishable by a number of phenotypic traits, such as flower colour (white and pink, respectively), sepal and seed capsule size and shape, leaf shape and many other traits (Baker 1951).

S. latifolia and *S. dioica* have similar karyotypes (Grabowska-Joachimiak & Joachimiak 2002; also see Fig. 1 in Filatov *et al.* 2009). Sex chromosomes are cytologically distinguishable in males, with the Y-chromosome being the largest and the X-chromosome the second largest in the genome (Armstrong & Filatov 2008). In contrast, another closely related *S. latifolia* relative, *S. diclinis*, that has previously been used for the analysis of Haldane's rule in plants (Brothers & Delph 2010; Demuth *et al.* 2014), has evolved neo-sex–chromosomes that form a quadrivalent in male meiosis (Howell *et al.* 2009). As this major sex chromosome rearrangement may play a role in species isolation, *S. diclinis* is not suitable for testing the large-X effect. In *S. latifolia* and *S. dioica* males, meiotic pairing of the X- and Y-chromosomes occurs only at the end of one arm in a relatively small pseudoautosomal region (PAR), while the rest of the sex chromosomes remain unpaired (Armstrong & Filatov 2008), though a previous QTL mapping study indicated the presence of a second PAR (Scotti & Delph 2006). Recombination is completely suppressed along most of the Y-chromosome length, which is also the case in mammals (Marshall Graves 2008). However, unlike mammals, the *Silene* Y-chromosome is mostly non-degenerate and contains thousands of expressed and apparently intact genes (Bergero & Charlesworth 2011; Chibalina & Filatov 2011; Muyle *et al.* 2012). Due to the recent evolution of sex chromosomes in *Silene*, the divergence between homologous X- and Y-linked genes is quite low (<15% silent sites, Chibalina & Filatov 2011). Thus, most genes on the X-chromosome are not hemizygous in males, and the dominance effects between sex chromosomes could be similar to those between autosomes. In this paper, we evaluate the contribution of the X-chromosome to interspecific gene flow in *Silene*. We demonstrate that despite the recent origin of sex chromosomes in this system and the minimal differences in dominance effects

between sex chromosomes and autosomes, the X-chromosome displays disproportionately large levels of interspecific differentiation.

Materials and Methods

Plant materials

Two *S. latifolia* and three *S. dioica* female plants were grown in a glasshouse (20°C, 15h lighting) from seed collected in the wild (Table 1). In addition to that, one *S. latifolia* female (Sa833d) and one hermaphroditic *S. vulgaris* individual (an outgroup) from previously published work were used (Chibalina & Filatov 2011).

RNA extraction and RNA-seq

RNA was extracted from actively growing shoots with flower buds from all of the plants. Each sample was extracted and analysed separately. Neither tissue nor RNA from different samples was pooled. Total RNA was extracted from all samples using a Qiagen RNeasy Plant Mini Kit with on-column DNAse digestion. Isolation of mRNA, cDNA synthesis and high-throughput sequencing were conducted according to the standard Illumina RNA-Seq procedure. High-throughput sequencing for each individual was conducted using an Illumina HiSeq instrument with 100 base, paired-end reads.

As a reference transcriptome we used 19195 *S. latifolia* cDNA contigs previously reported in Chibalina & Filatov (2011). Some features of these contigs (e.g., the length and homology to known proteins and the sex-linkage) can be found in the supplemental materials of Chibalina & Filatov (2011). This reference transcriptome was assembled from RNA extracted from a female *S. latifolia,* and thus any sex-linked contigs are necessarily X-linked (rather than a mixture of X- and Y-linked contigs).

Three previous studies have reported sex-linked genes in *S. latifolia* (Bergero & Charlesworth 2011; Chibalina & Filatov 2011; Muyle *et al.* 2012). To minimise the number of non-sex-linked genes misclassified as X-linked, we limited our set of X-linked genes to those identified in Chibalina & Filatov (2011). The latter study used a conservative approach to identify X-linked genes based on segregation analysis in two separate genetic crosses. Only the genes that segregated as X-linked in both crosses were classified as X-linked. Other reports of sex-linked genes have relied either on one cross (Bergero & Charlesworth 2011) or were based on identification of male-specific (presumably Y-linked) polymorphisms without specifically testing X-linkage (Muyle *et al.* 2012). Thus, to minimise erroneous inclusion of X-linked genes into the autosomal set, we i) excluded all the genes identified as sex-linked in the other studies, and ii) used the data from Chibalina & Filatov (2011) to identify and exclude genes that contained SNPs with a segregation pattern compatible with a sex-linkage pattern where inheritance was from the male parent to male progeny only or to female progeny only.

Single nucleotide polymorphism (SNP) calling

CLC Genomics Workbench (CLCgwb) (http://www.clcbio.com) was used to map RNA-seq reads to the reference transcriptome (Chibalina & Filatov 2011), with a maximum of two mismatches between the read and the reference and removal of the reads that were mapped to more than one reference position. The probabilistic SNP caller in CLCgwb was used to call SNPs with a variant probability ≥ 90%, a minimum SNP coverage of 10 reads, and at least 3 reads for each allele. Any SNPs with a quality score of less than 20 (phred scale) were discarded.

Gene flow analysis

To compile a DNA polymorphism dataset for gene flow analysis, we followed the same strategy used by Chapman *et al.* (2013). Briefly, the reference transcriptome and SNP calls for three *S. latifolia* and three *S. dioica* females from wild accessions (see above) were imported into a project database in ProSeq3 (Filatov 2009) and reformatted as sequence alignments. Information about the locations of coding regions available for the reference transcriptome (Chibalina & Filatov 2011) was used to extract 4-fold degenerate sites with ProSeq3. A previously published transcriptome from *S. vulgaris* (Chibalina & Filatov 2011) was used as outgroup. The alignments of 4-fold degenerate sites longer than 99 bases were used to calculate descriptive statistics, such as average pairwise nucleotide differences, π (Nei 1987, eq. 10.5) and population scaled mutation rate, θ (Watterson 1975, eq.1.4a). Total and net population divergence, D_{xy} and D_a (Nei 1987, eqs.10.20 and 10.21, respectively) and population differentiation, F_{st} (Hudson *et al.* 1992, eq.9) were calculated for genes with at least 5 polymorphic 4-fold degenerate sites.

To test for interspecific gene flow, we used an implementation of the isolation-with-migration model, IM (Hey 2010; Wang & Hey 2010) that was designed for analysis of very large numbers of sequenced loci sampled from a few individuals (Lohse *et al.* 2011). This approach is based on analytic likelihood computations that are tractable only for the samples of one to two individuals per species, which may be problematic if strong intra-species population structure results in different population histories for the different individuals from the same species. To circumvent this problem, we ran the analysis multiple times by using different individuals sampled from various places across the European species ranges (Table 1). Under this IM model, the overall distance between sequences sampled from two species (or populations) is dependent on (1) the time since species split (T) and (2) the time to coalescence in the ancestral species (τ) (Takahata 1986; Yang 2010). The latter accounts for incomplete lineage sorting (ILS). In the absence of gene flow, T is the same for all genes and variation in coalescence time is due entirely to variation of τ among the genes. Migration between the species introduces additional variance among the genes, as T is no longer the

same for different genes. Some genes may coalesce at a time shorter than *T* while other genes may coalesce at a time longer than *T* (ILS events). Both cases are taken into account by the IM model in estimating the migration parameter.

Using this approach, we obtained the maximum likelihood estimates (MLEs) of migration ($M = 4N_e m$), population diversity ($\theta = 4N_e\mu$), and speciation time (T), where N_e, m and μ stand for effective population size, migration rate, and mutation rate per generation, respectively. Estimates of parameter values were converted by assuming a generic per nucleotide per generation mutation rate of 5×10^{-9} (Muir *et al.* 2012). Significance of interspecific gene flow was tested using the likelihood ratio test (LRT) by comparing the nested models with and without migration between species. For significance testing we assumed that the difference between the log-likelihoods for nested models with and without gene flow is approximately χ^2-distributed with one degree of freedom. The analysis was conducted with *Mathematica* v9.01 (Wolfram Inc.) using the scripts of Lohse and co-workers (Lohse *et al.* 2011).

Results

To investigate the extent of species divergence and interspecific gene flow on the plant X-chromosome and the autosomes, we analyzed the transcriptome-wide DNA polymorphisms in the three individuals of each of *Silene latifolia* and *S. dioica*. The DNA polymorphism dataset consisted of 9158 autosomal and 659 X-linked genes that were previously identified by Chibalina & Filatov (2011). To minimise the effects of selection on demographic inferences, we limited all sequence polymorphism/divergence analyses to the 4-fold degenerate sites in the protein coding genes.

Levels of DNA polymorphism were very similar in the two species (Fig. 1A). The average pairwise nucleotide difference per-site within species (π) was significantly lower in X-linked genes (π_X) than in autosomal genes (π_A) in both *S. latifolia* ($\pi_X = 0.0088 \pm 0.00433$,

π_A = 0.0117 ± 0.0087; Wilcoxon rank sum test $P < 0.0001$) and *S. dioica* (π_X= 0.0093 ± 0.00429, π_A = 0.0122 ± 0.00907; Wilcoxon rank sum test $P < 0.0001$). The observed π_X/π_A ratios (π_X/π_A = 0.752 in *S. latifolia*, π_X/π_A = 0.762 in *S. dioica*) are very close to the expectation of π_X/π_A = 0.75 from the ploidy difference between X-linked and autosomal genes.

Population differentiation (F_{st}) and net sequence divergence (D_a) between *S. latifolia* and *S. dioica* samples were significantly higher for the X-chromosome (median F_{st}=0.381, D_a= 0.00507) than for the autosomes (median F_{st}= 0.154, D_a= 0.00265; Wilcoxon rank sum test $P = 2.62 \times 10^{-10}$ for F_{st} and $P = 5.82 \times 10^{-5}$ for D_a). The correlation between F_{st} and D_a was highly significant ($P < 10^{-10}$) for X-linked and autosomal genes (r = 0.507 and 0.598, respectively); 25% and 36% of overall variation in F_{st} was determined by the net sequence divergence between the species for X-linked and autosomal genes, respectively. The F_{st} statistic is heavily affected by the level of polymorphism in populations. However, after excluding low polymorphism regions (with π < 0.1% in either species) that tend to generate higher F_{st} (Cruickshank & Hahn 2014), we still obtained significantly higher F_{st} values for the X-chromosome (median F_{st} = 0.271) than for the autosomes (median F_{st} = 0.135; Wilcoxon rank sum test $P < 0.0001$). A significant difference between the X-chromosome and autosomes was also detected for the average pairwise sequence divergence, D_{xy} (median D_{xy}= 0.0193 and 0.0147, Wilcoxon rank sum test W = 402097.5, $P = 1.006 \times 10^{-6}$). The X-linked genes showed much greater heterogeneity in the F_{st} distribution among sites in comparison with that for the autosomal genes (Fig. 1B,C), indicating that a non-negligible fraction of X-linked genes could be involved in species isolation. Taken together, these results indicate the existence of higher species differentiation and a lower amount of interspecific gene flow for X-linked genes compared with autosomal genes.

The maximum likelihood estimates (MLEs) of gene exchange between *S. latifolia* and *S. dioica* indicated substantially lower gene flow occurring on the X-chromosome: the estimated number of migrants per gene and generation is 1.75±0.38 and 0.64±0.26 for the autosomal ($M_a=4N_em_A$) and the X-linked ($M_X=3N_em_X$) genes, respectively (Table 2). The MLE of

population diversity ($\theta \sim N_e\mu$) estimated separately for the X-linked (per nucleotide θ_X=0.007±0.00085) and autosomal (per nucleotide θ_A=0.0099±0.00082) genes are fairly close to the estimates calculated directly from the average pairwise differences between sequences within species or populations (π). Both θ_X/θ_A and π_X/π_A ratios are close to the expectation of 75% from the ploidy difference between the X-chromosome and the autosomes. The results are robust to the choice of individuals used in the analysis (Table 2).

Likelihood ratio tests (LRTs) demonstrated that for the autosomal genes, gene flow between *S. latifolia* and *S. dioica* is significantly different from zero (Table 2). For the X-linked genes, however, none of the LRTs was significant, indicating that inclusion of the gene flow parameter does not significantly improve the fit of the model to data. The lack of significance in LRTs for X-linked genes may result from less data (fewer loci) being available for the X-chromosome, compared with the number of loci for the autosomes. To test this possibility we estimated the power of rejecting the no-gene-flow model with smaller subsets of the autosomal genes (Figure 1 and Suppl. Table S1). With subsets of only 100 autosomal genes, 60% of LRTs rejected the no-gene-flow model; while with subsets of 400 or more genes, all LRTs became significant ($P < 0.05$, Figure 1). This indicates that the number of 400 or more X-linked genes should be sufficient for detecting significant gene flow if the extent of gene flow on the X-chromosome was comparable to that on the autosomes. Thus, the lack of significance in LRT for the gene flow parameter in the X-linked genes cannot be explained by smaller sample size, but rather suggests a genuine reduction in *S. latifolia*-*S. dioica* gene flow for X-linked genes compared with inter-specific gene flow for autosomal genes.

Discussion

Previous analyses of DNA polymorphism in dioecious *Silene* examined only a few genes that showed an extremely wide range of polymorphism levels (Filatov *et al.* 2001;

Filatov *et al.* 2000; Ironside & Filatov 2005; Laporte *et al.* 2005; Qiu *et al.* 2010). Our study took advantage of a high-throughput transcriptome sequencing technique (RNA-seq) to collect genome-wide polymorphism and divergence data in *S. latifolia* and *S. dioica*. The choice of RNA-seq allowed us to benefit from a previously published RNA-seq based segregation analysis that classified *S. latifolia* genes into sex-linked and non-sex-linked genes (Chibalina & Filatov 2011). It further allowed us to focus on synonymous mutations at 4-fold degenerate codon positions in expressed genes and protein coding regions, which practically excludes the influence of selection on demographic inference, and also to generate SNP data less prone to downward bias in F_{st} estimates compared with markers having higher mutation rates, such as microsatellites (Wang 2015; Whitlock 2015).

Our analysis demonstrates that the genome-wide average of silent sequence diversity is around 1% in both *S. latifolia* and *S. dioica*, which is more than twofold lower than previously reported averages (2.88% and 2.15%, respectively, estimated across eight autosomal and seven X-linked genes, Qiu *et al.* 2010). The exact reason for this discrepancy is unclear, but it may arise from the unconscious use of more polymorphic genes in the earlier study (e.g., such genes provide more markers for genetic mapping), or from smaller sample sizes in our study. However, the latter should not be a problem because sampling a large number of genes from a few individuals is known to be far more informative than sampling a few genes from more individuals (Felsenstein 2006; Pluzhnikov & Donnelly 1996). In fact, sampling one or two individuals per species is considered sufficient to infer population dynamics and address many interesting population genetic questions (Li & Durbin 2011; Lohse *et al.* 2011).

The most interesting finding from our analysis is the much lower level of gene flow and higher species differentiation occurring between *S. latifolia* and *S. dioica* for X-linked genes relative to autosomal genes. Such reduced interspecific gene flow and/or inflated species divergence for X(or Z)-linked genes has previously been reported in mice (Payseur *et al.* 2004), *Drosophila* (Machado *et al.* 2002), *Heliconius* butterflies (Martin *et al.* 2013) and flycatchers (Ellegren *et al.* 2012), but not in any plant species. In animal species this

difference was suggested to reflect one of “two rules of speciation”, the large-X effect (Coyne & Orr 1989). In this study, we have demonstrated the same effect holds true in two closely related dioecious species of campions and below we discuss several possible explanations for the effect.

Our analysis revealed significantly higher population subdivision (F_{st}) and species divergence (D_a and D_{xy}) as well as reduced gene flow between *S. latifolia* and *S. dioica* in X-linked genes. Elevated F_{st} is suggested to occur often in regions of reduced intra-species diversity, such as pericentromeric regions, though D_{xy} is robust to this effect (Cruickshank & Hahn 2014). On the other hand, D_{xy} for X-linked genes could be disproportionately inflated if mutation rates are higher in females than in males. However, similar synonymous divergence of X- and Y-linked *S. latifolia* genes from those in the outgroup species, *Silene vulgaris,* indicates no difference in mutation rates between the sexes (Chibalina & Filatov 2011).

Higher F_{st} for X-linked genes is expected, given that the X chromosome exists as a single copy in males, causing a lower effective population size (N_e), stronger genetic drift and reduced genetic diversity of X-linked genes compared with autosomal genes. However, this effect is relatively weak with respect to the X/A difference in F_{st}, with a theoretical maximum of ~1.333 (from F_{st}~1/[1+4$N_e m$] for the autosomal genes and F_{st}~1/[1+3$N_e m$] for the X-linked genes), and the actual reduction in DNA polymorphism (π) on the X-chromosome is marginal - only about 20% in each *Silene* species. Thus, the difference in copy number between the X-chromosome and the autosomes and the slightly reduced diversity on the X-chromosome do not explain the observed ~2.5–fold higher F_{st} for X-linked relative to autosomal genes. Nor can they explain the much greater heterogeneity in F_{st} for X-linked genes (Fig. 1 B and C). Thus, it is unlikely that the higher species divergence and reduced gene flow for X-linked genes are entirely due to the X/A difference in effective population size. This suggests that non-neutral processes, such as divergent selection, may

act on X-linked genes to reduce genetic exchange and maintain species divergence despite ongoing natural hybridisation and homogenisation in other parts of the genome.

Perhaps the most obvious non-neutral explanation for the inflated differentiation and reduced interspecific gene flow on the X-chromosome is the presence of hybrid incompatibility factors. Such factors have been reported in some animals (Coyne & Orr 2004; Jiggins *et al.* 2001; Masly & Presgraves 2007; Presgraves & Orr 1998), and in *Silene* species a greater reduction in pollen viability has been reported in hybrids compared with intra-specific crosses (Brothers & Delph 2010). However, it remains unclear whether any ecologically important traits are controlled by the X-linked factors in *Silene*, although this might be expected given the size of the X-chromosome in these species (about 10% of the ~3Gb genome, e.g. Armstrong & Filatov 2008).

Unlike autosomes, sex-chromosomes are represented unequally in the two sexes, and hence they are expected to accumulate mutations favouring one sex but harming the other (Friberg *et al.* 2011; Rice *et al.* 2009). The evolution of mutations advantageous to one sex can promote the evolution of restorers to minimise the harmful effect of such mutations on the opposite sex (Burt & Trivers 2006). In interspecific hybrids, however, sexually antagonistic mutations may not be 'neutralised' by restorers, and thus will harm the disadvantaged sex more severely than within the species (Berenos *et al.* 2012), which potentially promotes speciation (Crespi & Nosil 2013). If sexually antagonistic mutations are present on the *S. latifolia* X-chromosome, they could potentially reduce inter-specific gene flow and increase species divergence for X-linked genes in *Silene*. In this respect, it is of interest that the presence of sexually antagonistic mutations on *S. latifolia* sex chromosomes has recently been detected (Delph *et al.* 2011; Qiu *et al.* 2013). In addition, sex-linked factors causing a female-biased sex ratio have been reported in many *S. latifolia* populations (Taylor 1994a). Interestingly, X-linked female-biasing factor(s) appear to be counter-balanced by species-specific Y-linked restorers, and the sex-ratio bias is significantly more severe in interspecific hybrids (*S. latifolia* × *S. dioica*) than within the species (Taylor 1994b).

However, the female-bias in hybrids should bring the effective population size (and gene flow) of X-linked genes closer to that of autosomes. Nonetheless, co-evolution of sex-linked sexually antagonistic mutations and their restorers could potentially impede interspecific introgression for X-linked genes.

Inversions can contribute to interspecific divergence (Lohse *et al.* 2015; Navarro & Barton 2003), and sexually antagonistic mutations may drive rapid fixation of sex-linked chromosomal inversions that suppress recombination between the X- and Y-chromosomes in males (Rice 1987a). If X-linked inversions were fixed after the *S. latifolia* / *S. dioica* species split, such inversions could have contributed to reducing gene flow of X-linked genes. An X-linked inversion distinguishing *S. latifolia* and *S. dioica* was detected by genetic mapping (Nicolas *et al.* 2005), though, cytologically, the karyotypes of these two species are indistinguishable (Filatov *et al.* 2009; Grabowska-Joachimiak & Joachimiak 2002) and likely to be mostly collinear.

Dosage compensation evolving on the X-chromosome to compensate for the gradual degeneration of the Y-linked genes could potentially be another cause for the apparent inability of the X-chromosome to cross the *S. latifolia*-*S. dioica* species boundary. Both significant transcriptional deterioration of Y-linked genes (Chibalina & Filatov 2011) and at least partial dosage compensation of X-linked genes (Muyle *et al.* 2012) have been reported in *S. latifolia.* If the dosage compensation is gene-specific, then different sets of genes may be dosage compensated in the two species. Bringing together the X-chromosome of one species and the Y from the other may result in a significant transcriptional imbalance of the sex-linked genes in hybrids. This mechanism of species incompatibility may not be particularly pronounced in species with ancient sex chromosomes, such as mammals and many *Drosophila* species, as only a few Y-linked genes remain functional and further genetic degeneration is very slow (e.g., mammalian Y-chromosome degeneration; Bellott *et al.* 2014). Therefore, reduced gene flow on the X may be more important for species with recently evolved sex chromosomes, where the loss of transcriptional activity in Y-linked

genes and dosage compensation in X-linked genes are progressing. This hypothesis deserves further thorough investigation.

Regardless of the exact mechanism causing reduced interspecific gene flow for X-linked genes relative to autosomal genes, a lower level of gene exchange should promote higher interspecific divergence on the X chromosome, and this was found to be the case in our study where net sequence divergence, D_a, was significantly higher on the X-chromosome. The overall low values estimated for D_a (<0.007) indicate a very recent divergence of the two species, which is consistent with our ML estimate of the species split time ($<0.5 \times 10^6$ generations). However, the latter estimate has to be treated with caution as it depends on the assumption that per nucleotide per generation mutation rate is $\sim 5 \times 10^{-9}$, while the real mutation rate in *Silene* is unknown. Furthermore, the IM model used in the analysis of gene flow and species divergence is quite simplistic and the real speciation scenario is likely to be more complex. In particular, our analysis assumed constant population size, which may not necessarily be the case. However, violation of the assumptions should not differentially affect autosomal and X-linked genes.

Reduced interspecific gene flow on the X-chromosome may have implications for previous studies that have compared levels of intraspecific polymorphism between sex-linked and autosomal genes in *S. latifolia* (Filatov *et al.* 2001; Qiu *et al.* 2010). These studies reported that DNA polymorphism on the *S. latifolia* Y-chromosome is dramatically reduced, compared to other chromosomes, as expected from population genetic models describing ongoing genetic degeneration of non-recombining Y-chromosomes (Charlesworth *et al.* 1995; Rice 1987b). Higher interspecific gene flow for autosomal relative to sex-linked genes is expected to inflate intra-species polymorphism on the autosomes. However, this effect is likely to be relatively weak, as sequence divergence between *S. latifolia* and *S. dioica* is low. Furthermore, the bulk of evidence for the reduced polymorphism on *S. latifolia* Y-chromosome comes from comparisons between X- and Y-linked genes (Filatov *et al.* 2000; Laporte *et al.* 2005; Qiu *et al.* 2010), which should be less affected by interspecific gene flow

because both Y- and X-chromosomes are not actively introgressing between the two species (Ironside & Filatov 2005 and this study).

Previous studies have addressed the role of sex chromosomes in speciation by focusing on animal species where X-linked genes are predominantly hemizygous, such as *Drosophila* (Masly & Presgraves 2007), mice (Dod *et al.* 1993) and birds (Ellegren *et al.* 2012), but see Presgraves and Orr (1998) for an example of a study in a species with non-hemizygous X-chromosome in males. The X (or Z)-chromosomes in such animal species often play a disproportionately large role in speciation. While many different mechanisms have been identified, most of these depend on hemizygosity of X-linked genes (Qvarnstrom & Bailey 2009). However, there are many species where the X-chromosome is not or is unlikely to be hemizygous. About 5% of all flowering plants are dioecious (Yampolsky & Yampolsky 1922), some of which are known to have sex chromosomes (Charlesworth 2008; Westergaard 1958). Most of the plant species where sex chromosomes (or fully sex-linked chromosomal regions) have been found to be present, are agriculturally important (e.g., hop and papaya), suggesting that many other less studied dioecious species may also contain sex chromosomes. Given that the plant sex chromosomes studied so far (papaya, *Rumex* and *Silene*) show relatively little genetic degeneration of Y-linked genes (Bergero & Charlesworth 2011; Chibalina & Filatov 2011; Hough *et al.* 2014; Muyle *et al.* 2012; Wang *et al.* 2012), and that extensive haploid expression in pollen appears to slow down Y-chromosome degeneration in plants (Chibalina & Filatov 2011), it appears likely that many plant X-linked genes are non-hemizygous in males. Thus, studies on species with non-degenerate Y-chromosomes provide an informative comparison for a better understanding of the causes of Haldane's rule and the large-X effect (Demuth *et al.* 2014; Presgraves & Orr 1998).

Our report of reduced interspecific gene flow of X-linked genes in *Silene* provides an opportunity for further studies on the role of sex chromosomes in speciation. Further studies will greatly benefit from a detailed genetic map and genomic resources of *Silene* species to enable pinpointing the causal factors/genes that are subject to diversifying selection, and

measurement of the distribution of species divergence in the proximity of such regions. It will also be informative to compare the relative effects of sex-chromosomes and autosomes on various morphological and ecological traits that distinguish *S. latifolia* and *S. dioica*. This could reveal the nature of diversifying selective pressure and the role of sex chromosomes in maintaining the distinctiveness of these closely related hybridising species.

Acknowledgements

We thank Konrad Lohse and Nick Barton for providing us Python scripts, and Alex Papadopoulos and Michael Chester for helpful suggestions and corrections to the manuscript. We are grateful to Konrad Lohse for a helpful discussion and useful suggestions on the implementation of the IM model. This work was supported by funding to DAF from NERC (NE/F018991/1) and BBSRC (BB/K016539/1).

References

Armstrong SJ, Filatov DA (2008) A cytogenetic view of sex chromosome evolution in plants. *Cytogenet. Genome Research* **120**, 241-246.

Baker HG (1948) Stages in invasion and replacement demonstrated by species of *Melandrium*. *Journal of Ecology* **36**, 96-119.

Baker HG (1951) The inheritance of certain characteristics in crosses between *Melandrium dioicum* and *M. album*. *Genetica* **25**, 126-156.

Bellott DW, Hughes JF, Skaletsky H*, et al.* (2014) Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494-499.

Berenos C, Schmid-Hempel P, Wegner KM (2012) Antagonistic coevolution accelerates the evolution of reproductive isolation in *Tribolium castaneum*. *American Naturalist* **180**, 520-528.

Bergero R, Charlesworth D (2011) Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Current Biology* **21**, 1470-1474.

Bergero R, Qiu S, Forrest A, Borthwick H, Charlesworth D (2013) Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. *Genetics* **194**, 673-686.

Bernasconi G, Antonovics J, Biere A*, et al.* (2009) *Silene* as a model system in ecology and evolution. *Heredity (Edinb)* **103**, 5-14.

Brothers AN, Delph LF (2010) Haldane's rule is extended to plants with sex chromosomes. *Evolution* **64**, 3643-3648.

Burt A, Trivers R (2006) *Genes in conflict : the biology of selfish genetic elements.* Belknap Press of Harvard University Press., Cambridge, Mass.

Chapman MA, Hiscock SJ, Filatov DA (2013) Genomic divergence during speciation driven by adaptation to altitude. *Molecular Biology and Evolution* **30**, 2553-2567.

Charlesworth B, Coyne JA, Barton NH (1987) The relative rates of evolution of sex-chromosomes and autosomes. *American Naturalist* **130**, 113-146.

Charlesworth D (2008) Plant sex chromosomes. *Genome Dynamics* **4**, 83-94.

Charlesworth D, Charlesworth B, Morgan MT (1995) The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619-1632.

Chibalina MV, Filatov DA (2011) Plant Y chromosome degeneration is retarded by haploid purifying selection. *Current Biology* **21**, 1475-1479.

Coyne JA, Orr HA (1989) Two rules of speciation. In: *Speciation and Its Consequences* (eds. Otte D, Endler J), pp. 180-270. Sinauer Associates, Sunderland, Maryland.

Coyne JA, Orr HA (2004) *Speciation.* Sinauer Associates, Sunderland, Maryland.

Crespi B, Nosil P (2013) Conflictual speciation: species formation via genomic conflict. *Trends in Ecology and Evolution* **28**, 48-57.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**, 3133-3157.

de Vienne DM, Hood ME, Giraud T (2009) Phylogenetic determinants of potential host shifts in fungal pathogens. *Journal of Evolutionary Biology* **22**, 2532-2541.

Delph LF, Andicoechea J, Steven JC*, et al.* (2011) Environment-dependent intralocus sexual conflict in a dioecious plant. *New Phytol* **192**, 542-552.

Demuth JP, Flanagan RJ, Delph LF (2014) Genetic architecture of isolation between two species of *Silene* with sex chromosomes and Haldane's rule. *Evolution* **68**, 332-342.

Dod B, Jermiin LS, Boursot P*, et al.* (1993) Counterselection on sex-chromosomes in the *Mus musculus* European hybrid zone. *Journal of Evolutionary Biology* **6**, 529-546.

Ellegren H, Parsch J (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics* **8**, 689-698.

Ellegren H, Smeds L, Burri R*, et al.* (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**, 756-760.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* **23**, 691-700.

Filatov DA (2009) Processing and population genetic analysis of multigenic datasets with ProSeq3 software. *Bioinformatics* **25**, 3189-3190.

Filatov DA, Howell EC, Groutides C, Armstrong SJ (2009) Recent spread of a retrotransposon in the *Silene latifolia* genome, apart from the Y chromosome. *Genetics* **181**, 811-817.

Filatov DA, Laporte V, Vitte C, Charlesworth D (2001) DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. *Molecular Biology and Evolution* **18**, 1442-1454.

Filatov DA, Moneger F, Negrutiu I, Charlesworth D (2000) Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**, 388-390.

Frank SH (1991) Divergence of meiotic drive-suppressors as an explanation for sex-biased hybrid sterility and inviability. *Evolution* **45**, 262-267.

Friberg U, Stewart AD, Rice WR (2011) Empirical evidence for son-killing X chromosomes and the operation of SA-zygotic drive. *PLoS One* **6**, e23508.

Grabowska-Joachimiak A, Joachimiak A (2002) C-banded karyotypes of two *Silene* species with heteromorphic sex chromosomes. *Genome* **45**, 243-252.

Grant S, Houben A, Vyskot B*, et al.* (1994) Genetics of sex determination in flowering plants. *Developmental Genetics* **15**, 214-230.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* **12**, 101-109.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.

Hough J, Hollister JD, Wang W, Barrett SC, 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.

Howell EC, Armstrong SJ, Filatov DA (2009) Evolution of neo-sex chromosomes in *Silene diclinis*. *Genetics* **182**, 1109-1115.

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**, 138-151.

Hurst LD, Pomiankowski A (1991) Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* **128**, 841-858.

Ironside JE, Filatov DA (2005) Extreme population structure and high interspecific divergence of the *Silene* Y chromosome. *Genetics* **171**, 705-713.

Jiggins CD, Linares M, Naisbit RE*, et al.* (2001) Sex-linked hybrid sterility in a butterfly. *Evolution* **55**, 1631-1638.

Laporte V, Filatov DA, Kamau E, Charlesworth D (2005) Indirect evidence from DNA sequence diversity for genetic degeneration of the Y-chromosome in dioecious

species of the plant *Silene*: the SlY4/SlX4 and DD44-X/DD44-Y gene pairs. *Journal of Evolutionary Biology* **18**, 337-347.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496.

Lohse K, Clarke M, Ritchie MG, Etges WJ (2015) Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution* **69**, 1178-1190.

Lohse K, Harrison RJ, Barton NH (2011) A general method for calculating likelihoods under the coalescent process. *Genetics* **189**, 977-987.

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution* **19**, 472-488.

Macholan M, Munclinger P, Sugerkova M*, et al.* (2007) Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution* **61**, 746-771.

Marshall Graves JA (2008) Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annual Reviews Genetics* **42**, 565-586.

Martin SH, Dasmahapatra KK, Nadeau NJ*, et al.* (2013) Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research* **23**, 1817-1828.

Masly JP, Presgraves DC (2007) High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biology* **5**, e243.

Minder AM, Rothenbuehler C, Widmer A (2007) Genetic structure of hybrid zones between *Silene latifolia* and *Silene dioica* (*Caryophyllaceae*): evidence for introgressive hybridization. *Molecular Ecology* **16**, 2504-2516.

Muir G, Bergero R, Charlesworth D, Filatov DA (2011) Does local adaptation cause high population differentiation of *Silene latifolia* Y chromosomes? *Evolution* **65**, 3368-3380.

Muir G, Dixon CJ, Harper AL, Filatov DA (2012) Dynamics of drift, gene flow, and selection during speciation in *Silene*. *Evolution* **66**, 1447-1458.

Muir G, Filatov D (2007) A selective sweep in the chloroplast DNA of dioecious *Silene* (section *Elisanthe*). *Genetics* **177**, 1239-1247.

Muyle A, Zemp N, Deschamps C*, et al.* (2012) Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol* **10**, e1001308.

Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. *Science* **300**, 321-324.

Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York.

Nicolas M, Marais G, Hykelova V*, et al.* (2005) A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol* **3**, e4.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution* **58**, 2064-2078.

Pluzhnikov A, Donnelly P (1996) Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* **144**, 1247-1262.

Prentice H, Malm J, Hathaway L (2008) Chloroplast DNA variation in the European herb *Silene dioica* (red campion): postglacial migration and interspecific introgression. *Plant. Syst. Evol.* **272**, 23-37.

Presgraves DC, Orr HA (1998) Haldane's rule in taxa lacking a hemizygous X. *Science* **282**, 952-954.

Qiu S, Bergero R, Charlesworth D (2013) Testing for the footprint of sexually antagonistic polymorphisms in the pseudoautosomal region of a plant sex chromosome pair. *Genetics* **194**, 663-672.

Qiu S, Bergero R, Forrest A, Kaiser VB, Charlesworth D (2010) Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes. *Proc Biol Sci* **277**, 3283-3290.

Qvarnstrom A, Bailey RI (2009) Speciation through evolution of sex-linked genes. *Heredity (Edinb)* **102**, 4-15.

Rautenberg A, Sloan DB, Alden V, Oxelman B (2012) Phylogenetic relationships of *Silene multinervia* and *Silene* section *Conoimorpha* (*Caryophyllaceae*). *Systematic Botany* **37**, 226-237.

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 WR, Gavrilets S, Friberg U (2009) Sexually antagonistic chromosomal cuckoos. *Biol Lett* **5**, 686-688.

Scotti I, Delph LF (2006) Selective trade-offs and sex-chromosome evolution in *Silene latifolia*. *Evolution* **60**, 1793-1800.

Shull GH (1910) Inheritance of sex in Lychnis. *Bot. Gaz.* **52**, 110-125.

Takahata N (1986) An attempt to estimate the effective size of the ancestral species common to two extant species from which homologous genes are sequenced. *Genetical Research* **48**, 187-190.

Tao Y, Hartl DL (2003) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. III. Heterogeneous accumulation of hybrid incompatibilities, degree of dominance, and implications for Haldane's rule. *Evolution* **57**, 2580-2598.

Taylor DR (1994a) The genetic basis of sex ratio in *Silene alba* (= *S. latifolia*). *Genetics* **136**, 641-651.

Taylor DR (1994b) Sex ratio in hybrids between *Silene alba* and *Silene dioica*: evidence for Y-linked restorers. *Heredity (Edinb)* **73 (Pt 5)**, 518-526.

Teeter KC, Payseur BA, Harris LW*, et al.* (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research* **18**, 67-76.

Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM (1992) Abrupt cline for sex-chromosomes in a hybrid zone between two species of mice. *Evolution* **46**, 1146-1163.

Turelli M, Moyle LC (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* **176**, 1059-1088.

Turelli M, Orr HA (1995) The dominance theory of Haldane's rule. *Genetics* **140**, 389-402.

Turelli M, Orr HA (2000) Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* **154**, 1663-1679.

van Nigtevecht G (1966a) Genetic studies in dioecious *Melandrium* .2. Sex determination in *Melandrium album* and *Melandrium dioicum*. *Genetica* **37**, 307-344.

van Nigtevecht G (1966b) Genetic studies in dioecious *Melandrium* .I. Sex-Linked and sex-influenced inheritance in *Melandrium album* and *Melandrium dioicum*. *Genetica* **37**, 281-306.

Vicoso B, Charlesworth B (2006) Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews Genetics* **7**, 645-653.

Waelti MO, Muhlemann JK, Widmer A, Schiestl FP (2008) Floral odour and reproductive isolation in two species of *Silene*. *Journal of Evolutionary Biology* **21**, 111-121.

Wang J (2015) Does GST underestimate genetic differentiation from marker data? *Molecular Ecology* **24**, 3546-3558.

Wang J, Na JK, Yu Q*, et al.* (2012) Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci U S A* **109**, 13710-13715.

Wang Y, Hey J (2010) Estimating divergence parameters with small samples from a large number of loci. *Genetics* **184**, 363-379.

Watterson G (1975) On the number of segregating sites in genetic models without recombination. *Theoretical Population Biology* **7**, 256-276.

Westergaard M (1958) The mechanism of sex determination in dioecious flowering plants. *Adv Genet* **9**, 217-281.

Whitlock MC (2015) A clever solution to a vexing problem. *Molecular Ecology* **24**, 3513-3514.

Yampolsky C, Yampolsky H (1922) Distribution of sex forms in the phanerogamic flora. *Bibl. Genet.* **3**, 1-62.

Yang ZH (2010) A likelihood ratio test of speciation with gene flow using genomic sequence data. *Genome Biology and Evolution* **2**, 200-211.

Table 1. Samples of *Silene* species analysed in this study.

Individual	Species	Country	Approx. Location	Coordinates	
Sa526b	*S. latifolia*	Austria	Stift-Melk	15° 20.234' E	48° 13.762' N
Sa985a	*S. latifolia*	Austria	near Klagenfurt	14° 23.444' E	46° 32.610' N
Sa833d	*S. latifolia*	Spain	Broto	0° 7.103' W	42° 35.989' N
Sd43a	*S. dioica*	Wales	Talybont-on-Usk	3° 17.388' W	51° 53.512' N
Sd468c	*S. dioica*	Austria	St Oswald	15° 8.574' E	46° 42.410' N
Sd496a	*S. dioica*	Austria	B28 road	15° 17.685' E	47° 54.057' N
Sv	*S. vulgaris*	UK	near Oxford	1° 14.518' W	51° 52.263' N

Table 2. Maximum likelihood estimates of the parameters of the IM model for *S. latifolia* and *S. dioica*.

S.latifolia_S.dioica	*X-linked*				*Autosomal genes*			
sample pairs	*M* [a)]	*T* [a)]	θ [a)]	*LogL*	*M* [a)]	*T* [a)]	θ [a)]	*LogL*
Sa833d_Sd496a		0.551	1.947	-623.501		0.518	2.167	-19197.5
	0.722	0.685	1.877	-623.357	1.118	0.749	2.038	-19181.8***
Sa833d_Sd43a		0.383	2.205	-631.345		0.493	2.139	-19038.7
	0.832	0.439	2.174	-631.325	1.437	0.815	1.973	-19021.1***
Sa833d_Sd468c		0.663	1.713	-602.645		0.501	2.072	-18849
	0.447	0.775	1.664	-602.596	1.447	0.860	1.893	-18825.3***
Sa985a_Sd496a		0.408	2.256	-600.974		0.416	2.411	-19155.2
	0.414	0.438	2.239	-600.961	1.994	0.747	2.233	-19134.3***
Sa985a_Sd43a		0.395	2.163	-590.031		0.399	2.390	-19034.9
	0.548	0.432	2.143	-590.022	2.175	0.750	2.208	-19016.7***
Sa985a_Sd468c		0.573	1.862	-574.607		0.397	2.324	-18834.3
	0.308	0.620	1.840	-574.583	2.252	0.801	2.123	-18810.1***
Sa526b_Sd496a		0.443	2.357	-652.312		0.440	2.548	-19653.3
	0.984	0.557	2.286	-652.158	1.568	0.674	2.406	-19638.6***
Sa526b_Sd43a		0.358	2.390	-652.301		0.430	2.534	-19584.8
	1.044	0.427	2.348	-652.290	1.815	0.739	2.353	-19563.6***
Sa526b_Sd468c		0.519	2.071	-627.817		0.420	2.499	-19461.3
	0.506	0.591	2.032	-627.75	1.980	0.781	2.296	-19436.1***
Mean [b)]	0.645	0.552	2.067		1.754	0.768	2.169	
SD [b)]	0.262	0.127	0.231		0.383	0.053	0.175	

*** $P < 0.0001$; Significance of gene flow was tested in likelihood ratio tests (LRTs) comparing a pair of nested models – with and without gene flow (bottom and top lines for each sample pair).

[a)] Estimated parameters of the IM model (Lohse *et al.* 2011): number of migrants per generation ($M = 4N_em$), the speciation time (T) and the population-scaled mutation rate (aka population diversity, $\theta = 4N_e\mu$), where N_e, m and μ stand for effective population size, migration and mutation rates per generation, respectively. Note that θ values in this table are calculated per gene, while the θ values shown in Results section are per nucleotide (i.e. divided by gene length).

[b)] Only the estimates from the most parameter rich model (bottom line in each sample pair) were used for calculating the mean and standard deviation (SD).

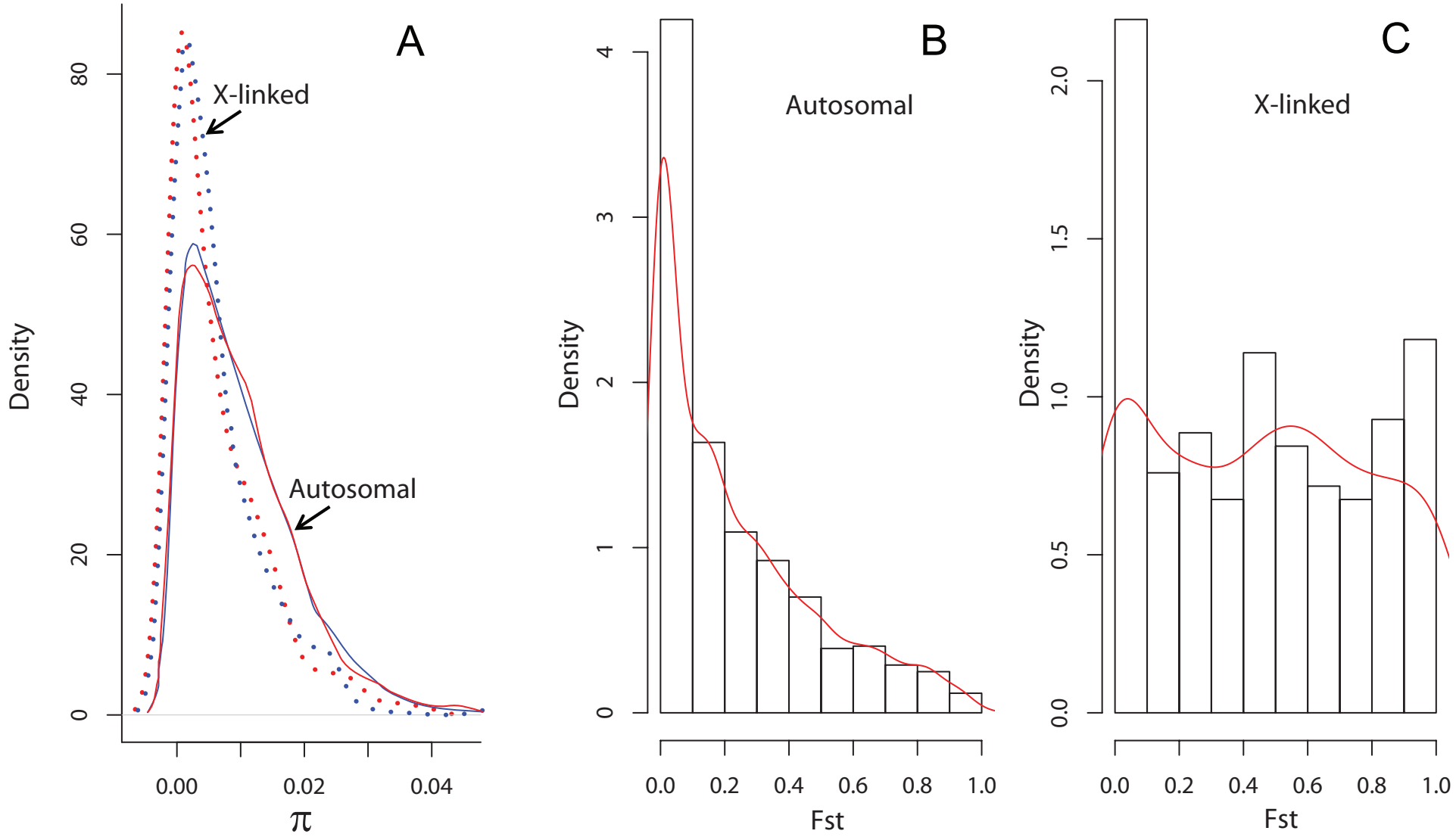
Figure legends

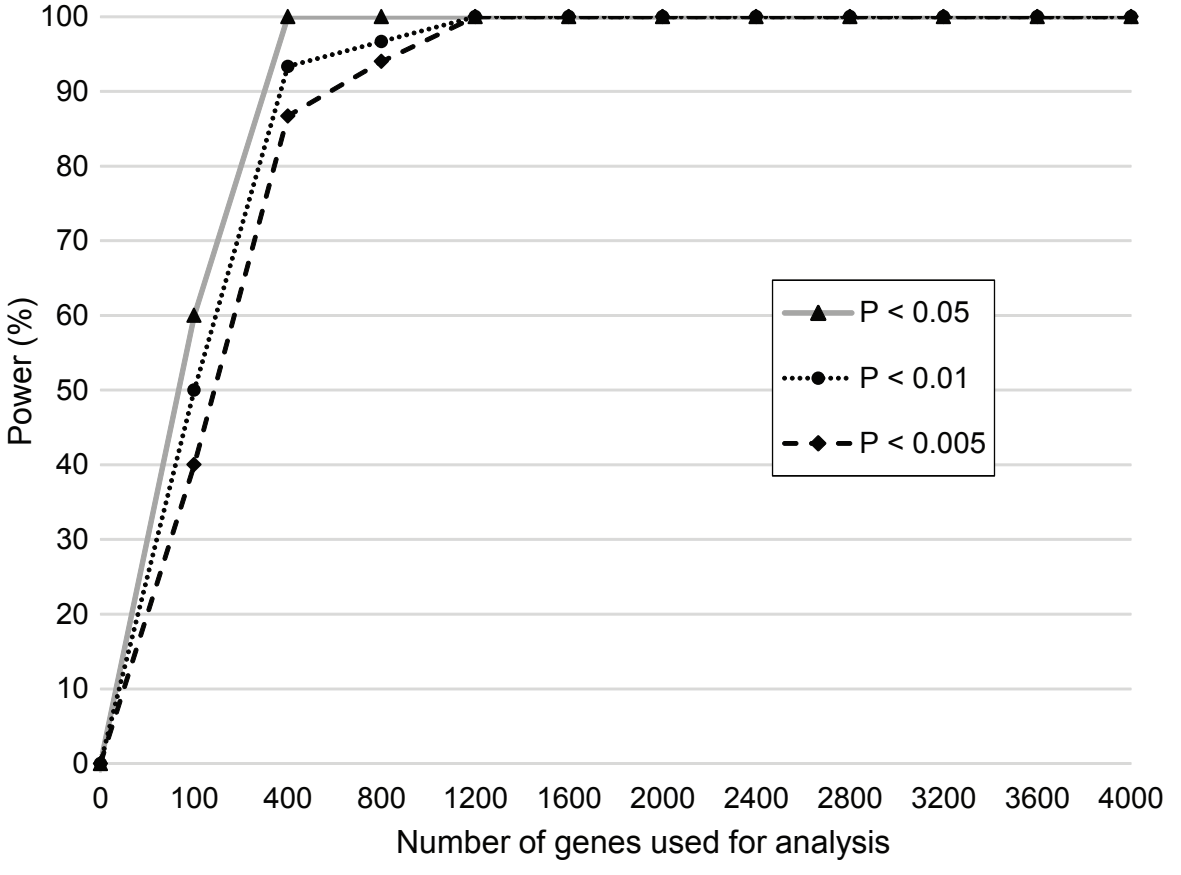
Figure 1. Distributions of nucleotide polymorphism and species differentiation across X-linked and autosomal genes in *S. latifolia* and *S. dioica*. A) Average pairwise sequence diversity per nucleotide (π) in X-linked (dotted line) and autosomal (solid line) genes in *S. latifolia* (blue) and *S. dioica* (red). B,C) Probability density histograms for species differentiation (F_{st}) in autosomal (B) and X-linked (C) genes.

Figure 2. Power to reject no-gene-flow model as a function of number of genes used in the analysis. The vertical axis shows the proportion of random subsets of genes that allowed rejection of the no-gene-flow model in the IM analysis (Lohse *et al.* 2011). The horizontal axis shows the size of the subsets of genes used in each of the analyses. See suppl. Table S1 for results of individuals runs with the subsets of genes.

Supplementary information

Suppl. Table S1: The ML parameter estimates and the power to reject the no-gene-flow model with random sub-sets of genes of a given size.


A
X-linked
Autosomal
Density
0
20
40
60
80
0.00
0.02
0.04
π
B
Autosomal
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Fst
C
X-linked
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
Fst


100
90
80
70
60
50
40
30
20
10
0
Power (%)
0
100
400
800
1200
1600
2000
2400
2800
3200
3600
4000
Number of genes used for analysis
P < 0.05
P < 0.01
P < 0.005